



US011136559B2

(12) **United States Patent**  
**Maeda et al.**

(10) **Patent No.:** **US 11,136,559 B2**  
(45) **Date of Patent:** **Oct. 5, 2021**

(54) **ENGINEERED PREPHENATE  
DEHYDROGENASES AND AROGENATE  
DEHYDROGENASES AND METHODS OF  
USING THE SAME**

(71) Applicant: **Wisconsin Alumni Research  
Foundation**, Madison, WI (US)

(72) Inventors: **Hiroshi A. Maeda**, Madison, WI (US);  
**Craig Albert Schenck**, Madison, WI  
(US)

(73) Assignee: **WISCONSIN ALUMNI RESEARCH  
FOUNDATION**, Madison, WI (US)

(\* ) Notice: Subject to any disclaimer, the term of this  
patent is extended or adjusted under 35  
U.S.C. 154(b) by 0 days.

(21) Appl. No.: **15/882,636**

(22) Filed: **Jan. 29, 2018**

(65) **Prior Publication Data**  
US 2018/0216083 A1 Aug. 2, 2018

**Related U.S. Application Data**

(60) Provisional application No. 62/451,124, filed on Jan.  
27, 2017.

(51) **Int. Cl.**  
**C12N 15/82** (2006.01)  
**C12N 9/02** (2006.01)

(52) **U.S. Cl.**  
CPC ..... **C12N 9/001** (2013.01); **C12N 15/8251**  
(2013.01); **C12Y 103/01012** (2013.01)

(58) **Field of Classification Search**  
None  
See application file for complete search history.

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

7,396,664 B2 7/2008 Daly et al.  
9,701,977 B2 7/2017 Maeda et al.  
2015/0150157 A1\* 5/2015 Maeda ..... C12N 15/8251  
800/278  
2018/0265880 A1 9/2018 Maeda et al.

**FOREIGN PATENT DOCUMENTS**

WO WO 2002/090523 11/2002

**OTHER PUBLICATIONS**

Medicago truncatula arogenate dehydrogenase with a GenBank  
accession No. G7K796, published Jan. 25, 2012.\*  
Arachis hypogaea chloroplastic arogenate dehydrogenase 1, GenBank  
accession No. XP\_025622852, published May 24, 2019.\*  
Arachis ipaensis chloroplastic arogenate dehydrogenase 1, GenBank  
accession No. XP\_016180356, published Apr. 22, 2016.\*  
Schenck et al., 2015, Non-plastidic, tyrosine-insensitive prephenate  
dehydrogenases from legumes, Nature Chemical Biology 11: 52-57,  
published Nov. 17, 2014.\*

Keskin et al., 2004, A new, structurally nonredundant, diverse data  
set of protein-protein interfaces and its implications Protein Science  
13: 1043-1055.\*

Guo et al., 2004, Protein tolerance to random amino acid change,  
Proceedings of the National Academy of Sciences USA 101:  
9205-9210.\*

Thornton et al., 2000, From structure to function: Approaches and  
limitations, Nature Structural Biology, structural genomic supple-  
ment, Nov. 2000: 991-994.\*

Medicago truncatula prephenate/arogenate dehydrogenase, GenBank  
accession No. AES71257, published on Jun. 18, 2014.\*

Medicago truncatula prephenate/arogenate dehydrogenase, UniProt  
accession No. G7J2E9, published on Jan. 25, 2012.\*

Bonner et al., 2008, Cohesion Group Approach for Evolutionary  
Analysis of TyrA, a Protein Family with Wide-Ranging Substrate  
Specificities, Microbiology and Molecular Biology Reviews 72:  
13-53.\*

Ochrobactrum intermedium LMG 3301 prephenate/arogenate  
dehydrogenase, UniProt accession No. C4WNJ2, published Jul. 28,  
2009.\*

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

Ambawat, S. et al., MYB transcription factor genes as regulators for  
plant responses: An overview. Physiol. Mol. Biol. Plants 19, 307-  
321 (2013).

Beaudoin, G. A. W. & Facchini, P. J. Benzylisoquinoline alkaloid  
biosynthesis in opium poppy. Planta 240, 19-32 (2014).

Bentley, R. (University of S. The Shikimate Pathway—A Metabolic  
Tree with Many Branches. Crit Rev. Biochem Mol Biol 25, 307-84  
(1990).

Bonner, C. A. et al., Cohesion group approach for evolutionary  
analysis of TyrA, a protein family with wide-ranging substrate  
specificities. Microbiol. Mol. Biol. Rev. MMBR 72, 13-53 (2008).  
Bonner, C. A. et al., Distinctive enzymes of aromatic amino acid  
biosynthesis that are highly conserved in land plants are also present  
in the chlorophyte alga *Chlorella sorokiniana*. Plant Cell Physiol.  
36, 1013-1022 (1995).

Bonvin, J. et al., Biochemical characterization of prephenate  
dehydrogenase from the hyperthermophilic bacterium *Aquifex  
aeolicus*. Protein Sci. 15, 1417-32 (2006).

Brockington, S.F., et al., Phylogeny of the Caryophyllales sensu  
lato: Revisiting hypotheses on pollination biology and perianth  
differentiation in the core Caryophyllales. 2009 International Jour-  
nal of Plant Sciences 170: 627-643.

Brockington, S.F., et al., Lineage-specific gene radiations underlie  
the evolution of novel betalain pigmentation in Caryophyllales.  
New Phytol. 207, 1170-1180 (2015).

(Continued)

*Primary Examiner* — Bratislav Stankovic  
(74) *Attorney, Agent, or Firm* — Quarles & Brady LLP

(57) **ABSTRACT**

The invention generally relates to engineered prephenate  
dehydrogenases and arogenate dehydrogenases and methods  
of using the same. More specifically, the invention relates in  
part to compositions including engineered prephenate dehy-  
drogenases (PDH) polypeptides and engineered arogenate  
dehydrogenase (ADH) polypeptides with altered substrate  
preferences and tyrosine sensitivities and methods of using  
the same.

**11 Claims, 26 Drawing Sheets**  
**(21 of 26 Drawing Sheet(s) Filed in Color)**  
**Specification includes a Sequence Listing.**

(56)

## References Cited

## OTHER PUBLICATIONS

- Brockington, S.F., et al., Complex pigment evolution in the Caryophyllales. *New Phytol.* 190, 854-864 (2011).
- Cardoso, D. et al. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. *Am. J. Bot.* 99, 1991-2013 (2012).
- Chávez-Béjar, M. I. et al. Metabolic engineering of *Escherichia coli* for L-tyrosine production by expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from *Zymomonas mobilis*. *Appl. Environ. Microbiol.* 74, 3284-3290 (2008).
- Christendat, D., et al., Use of site-directed mutagenesis to identify residues specific for each reaction catalyzed by chorismate mutase-prephenate dehydrogenase from *Escherichia coli*. *Biochemistry (Mosc.)* 37, 1573-1580 (1998).
- Christendat, D. & Turnbull, J. L. Identifying groups involved in the binding of prephenate to prephenate dehydrogenase from *Escherichia coli*. *Biochemistry (Mosc.)* 38, 4782-4793 (1999).
- Connelly, J. A. & Conn, E. E. Tyrosine biosynthesis in *Sorghum bicolor*: isolation and regulatory properties of arogenate dehydrogenase. *Verlag der Zeitschrift für Naturforsch.* 41c,69-78 (1986).
- Dal Cin, V. et al. Identification of Genes in the Phenylalanine Metabolic Pathway by Ectopic Expression of a MYB Transcription Factor in Tomato Fruit. *Plant Cell* 23, 2738-2753 (2011).
- Delaux, P.M. et al., Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *2014 PLoS Genetics* 10: e1004487.
- Des Marais, D. L. To betalains and back again: A tale of two pigments. *New Phytol.* 207, 939-941 (2015).
- Dornfeld, C. et al. Phylobiochemical characterization of class-Ib aspartate/prephenate aminotransferases reveals evolution of the plant arogenate phenylalanine pathway. *Plant Cell* 26, 3101-3114 (2014).
- Facchini, P. J. Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 29-66 (2001).
- Fischer, R. S. & Jensen, R. A. Prephenate dehydrogenase (monofunctional). *Methods Enzymol.* 142, 503-507 (1987).
- Gaines, C. G., et al., L-Tyrosine regulation and biosynthesis via arogenate dehydrogenase in suspension-cultured cells of *Nicotiana glauca* Speg. et Comes. *Planta* 156,233-240 (1982).
- Galanie, S. et al., Complete biosynthesis of opioids in yeast. DOI: 10.1126/science.aac9373, Published Online Aug. 13, 2015.
- Gamborg, O.L. & Keeley, F. W. Aromatic metabolism in plants I. A study of the prephenate dehydrogenase from bean plants. *Biochim. Biophys. Acta* 115,65-72 (1966).
- Greenberg, A.K. & Donoghue, M.J. Molecular systematics and character evolution in Caryophyllaceae. *2011 Taxon* 60: 1637-1652.
- Hagel, J. M. & Facchini, P. J. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647-672 (2013).
- Hatlestad, G. J. et al., The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nat. Genet.* 47, 92-6 (2015).
- Hernández-Ledesma, P., et al. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. *Willdenowia* 45, 281-383 (2015).
- Hudson, G.S. et al., Chorismate mutase/prephenate dehydrogenase from *Escherichia coli* K12: purification, characterization, and identification of a reactive cysteine. *Biochemistry* 23, 6240-6249 (1984).
- Hunter, S. C. & Cahoon, E. B. Enhancing vitamin E in oilseeds: unraveling tocopherol and tocotrienol biosynthesis. *Lipids* 42, 97-108 (2007).
- Jensen, R. A. & Pierson, D. L. Evolutionary implications of different types of microbial enzymology for L-tyrosine biosynthesis. *Nature* 254, 667-671 (1975).
- Jung, E., et al., Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. *Proc. Natl. Acad. Sci. U. S. A.* 83, 7231-7235 (1986).
- Kristensen, C. et al., Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. *2015 Proceedings of the National Academy of Sciences of the United States of America* 102: 1779-1784.
- Kutchan, T. M. Alkaloid biosynthesis: The basis for metabolic engineering of medicinal plants. *Plant Cell* 7, 1059-1070 (1995).
- Legrand, P. et al., Biochemical Characterization and Crystal Structure of Synechocystis Arogenate Dehydrogenase Provide Insights into Catalytic Reaction. *2006 Structure* 14: 767-776.
- Leuchtenberger, W., et al., Biotechnological production of amino acids and derivatives: current status and prospects. *Appl. Microbiol. Biotechnol.* 69, 1-8 (2005).
- Lopez-Nieves, S. et al. "Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in caryophyllales." (2017) *New Phytologist*.
- Lütke-Eversloh, T. & Stephanopoulos, G. Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of *Escherichia coli*: generation and characterization of tyrosine-insensitive mutants. *Appl. Environ. Microbiol.* 71, 7224-7228 (2005).
- Lütke-Eversloh, T. et al., Perspectives of biotechnological production of L-tyrosine and its applications. *Appl. Microbiol. Biotechnol.* 77, 751-762 (2007).
- Maeda, H. et al. RNAi suppression of arogenate dehydratase reveals that phenylalanine is synthesized predominantly via the arogenate pathway in petunia petals. *Plant Cell* 22, 832-49 (2010).
- Maeda, H. & Dudareva, N. The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annu. Rev. Plant Biol.* 63, 73-105 (2012).
- Maeda, H., et al. Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. *Nat. Chem. Biol.* 7, 19-22 (2011).
- Millgate, A. G. et al. Analgesia: Morphine-pathway block in topi poppies. *Nature* 431, 413-414 (2004).
- Raman, S., et al., Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci. U. S. A.* 111, 17803-17808 (2014).
- Reyes-Prieto, A. & Moustafa, A. Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. *Sci. Rep.* 2, 955 (2012).
- Rippert, P., et al., Engineering plant shikimate pathway for production of tocotrienol and improving herbicide resistance. *Plant Physiol.* 134, 92-100 (2004).
- Rippert, P. & Matringe, M. Molecular and biochemical characterization of an *Arabidopsis thaliana* arogenate dehydrogenase with two. *Plant Mol. Biol.* 48, 361-368 (2002).
- Rippert, P. & Matringe, M. Purification and kinetic analysis of the two recombinant arogenate dehydrogenase isoforms of *Arabidopsis thaliana* 2002 *European Journal of Biochemistry* 269: 4753-4761.
- Rippert, P., et al., Tyrosine and Phenylalanine are Synthesized within the Plastids in *Arabidopsis*. *Plant Physiol.* 149, 1251-1260 (2009).
- Rubin, J. L. et al., "Enzymology of L-Tyrosine Biosynthesis in Mung Bean (*Vigna radiata* [L.] Wilczek)," 1979 *Plant Physiol.* 64:727-734.
- Sariaslani, F. S. Development of a combined biological and chemical process for production of industrial aromatics from renewable resources. *Annu. Rev. Microbiol.* 61, 51-69 (2007).
- Schenck, C.A., et al. Tyrosine biosynthesis, metabolism, and catabolism in plants. *Phytochemistry*. May 2018;149:82-102.
- Schenck, C.A., et al. "Conserved molecular mechanism of TyrA dehydrogenase substrate specificity underlying alternative tyrosine biosynthetic pathways in plants and microbes." (2017) *Frontiers in Molecular Biosciences*.
- Schenck, C.A., et al., Molecular Basis of the evolution of alternative tyrosine biosynthetic routes in plants. *2017 Nature Chemical Biology* 13: 1029-1035.
- Schenck, C.A., et al. Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nature Chemical Biology*, 2015, pp. 52-57, vol. 11.

(56)

**References Cited**

## OTHER PUBLICATIONS

Siehl, D.L. "The Biosynthesis of Tryptophan, Tyrosine, and Phenylalanine from Chorismate," 1999 Plant Amino Acids: Biochemistry and Biotechnology, Edited by Bijay K. Singh, pp. 171-204.

Song, J., et al., The TyrA family of aromatic-pathway dehydrogenases in phylogenetic context. BMC Biol. 3, 13 (2005).

Sun, W., et al. The Crystal Structure of Aquifex aeolicus Prephenate Dehydrogenase Reveals the Mode of Tyrosine Inhibition. J. Biol. Chem. 284, 13223-13232 (2009).

Sun, W., et al. Crystal Structure of Prephenate Dehydrogenase from Aquifex aeolicus. Insights into the Catalytic Mechanism. 2006 Journal of Biological Chemistry 281: 12919-12928.

Tatiersall, D.B., et al., Resistance to an herbivore through engineered cyanogenic glucoside synthesis. 2001 Science 293: 1826-1828.

Vannelli, T., et al., Production of p-hydroxycinnamic acid from glucose in *Saccharomyces cerevisiae* and *Escherichia coli* by expression of heterologous genes from plants and fungi. Metab. Eng. 9, 142-151 (2007).

Wang, Y., et al., Metabolic engineering of flavonoids in plants and microorganisms. Appl. Microbiol. Biotechnol. 91, 949-956 (2011).

Wierenga, R. K., et al. Interaction of pyrophosphate moieties with alpha-helices in dinucleotide-binding proteins. Biochemistry (Mosc.) 24, 1346-1357 (1985).

Wojciechowski, M. F., et al., A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. Am. J. Bot. 91, 1846-62 (2004).

Yang, Y., et al. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. 2015 Molecular Biology and Evolution, 32, 2001-2014.

Office Action for U.S. Appl. No. 14/548,216, dated Mar. 28, 2016.

Office Action for U.S. Appl. No. 14/548,216, dated Nov. 25, 2016.

\* cited by examiner

Fig. 1A

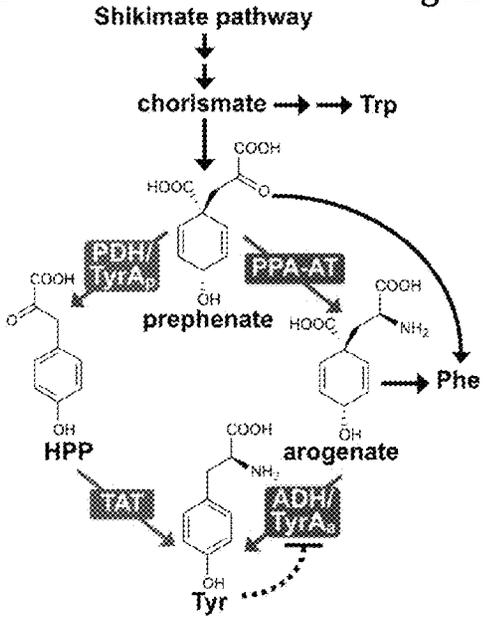


Fig. 1B

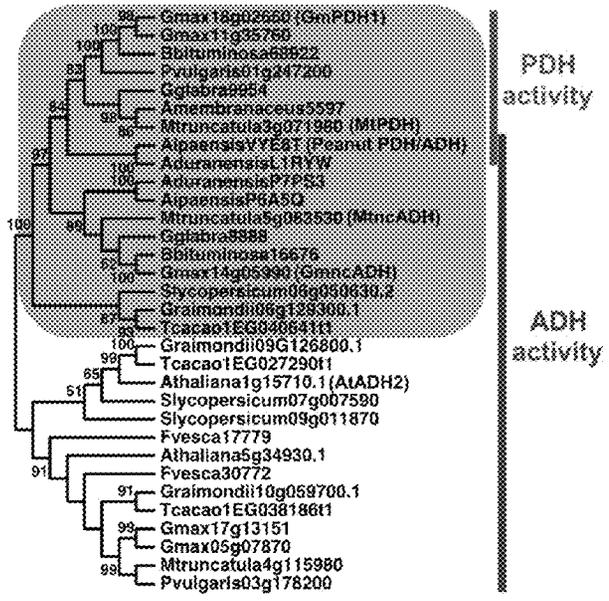


Fig. 1C

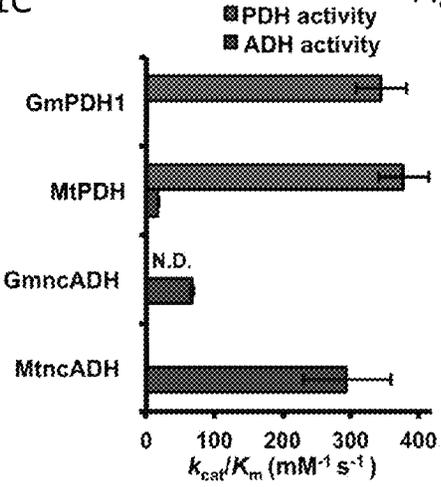
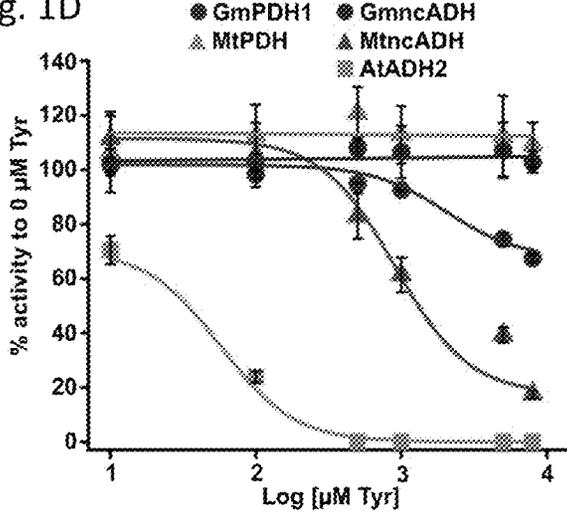


Fig. 1D



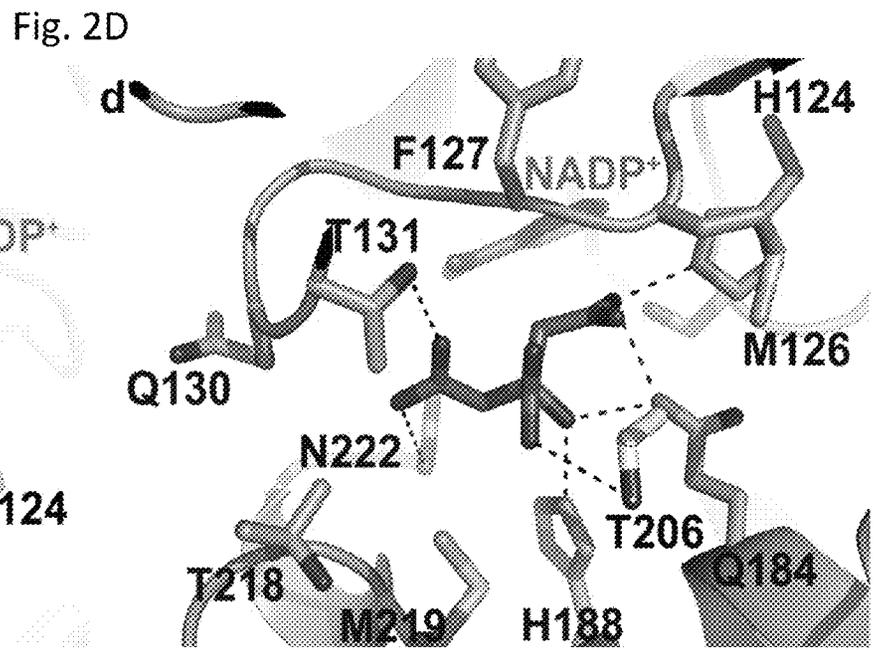
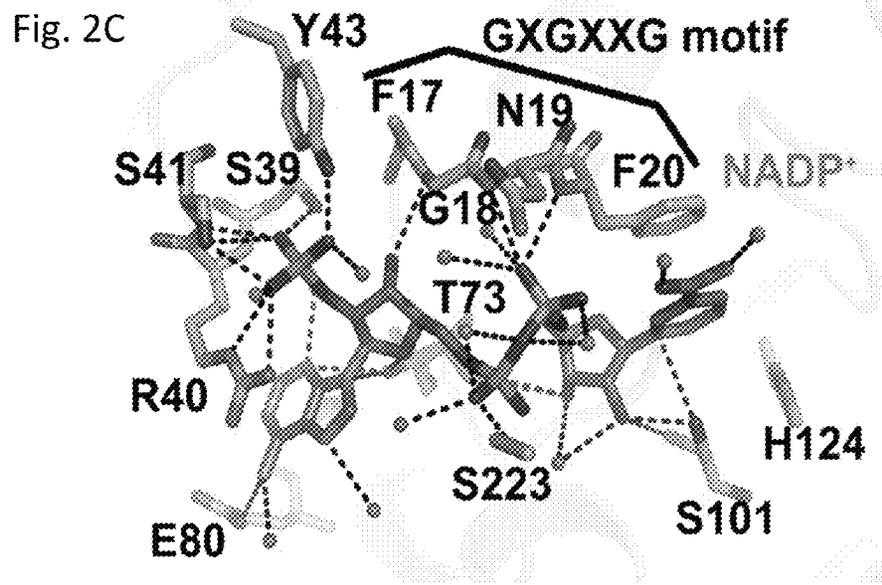
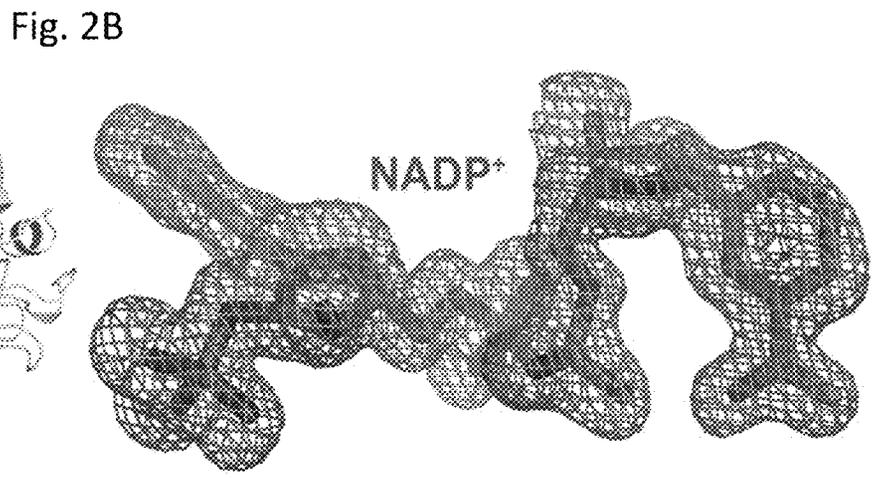
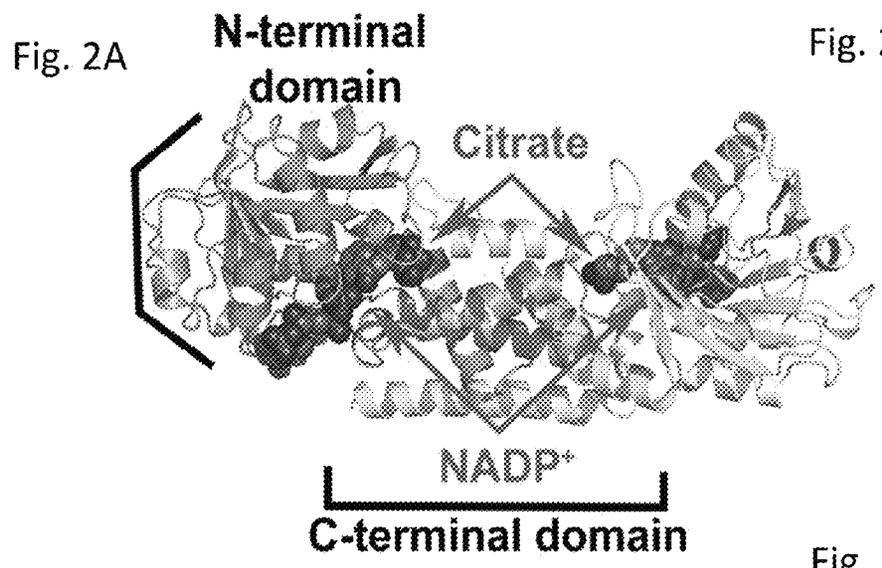
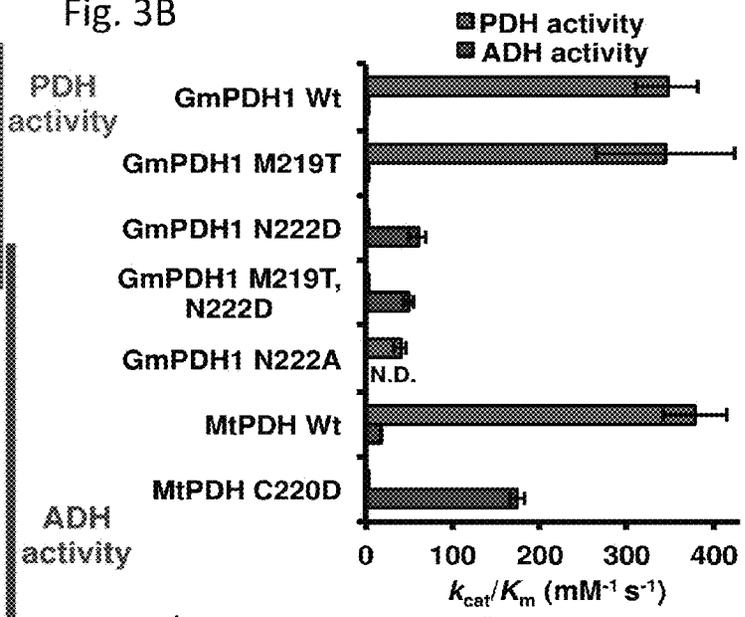


Fig. 3A

	M219	N222	
GmPDH1	KLKE*	M RNSFDLY	PDH activity
Gmax11g35760	KLKE*	M RNSFDLY	
Bbituminosa68922	KLKE*	M KNSFDLY	
Pvulgaris01g247200	KLKE*	M GNSFDLY	
Gglabra9954	IKKE*	V GCSSFDLY	
Amembranaceus5597	NILKE*	V GCSSFDLY	
MtPDH	ELKKP*	V GCSSFDLY	
Peanut PDH/ADH	ILKE*	T KCSFDLY	
AduranensisL1RYW	RYIKKE*	T KCSFDLY	
AduranensisP7PS3	ILKE*	T RDSFDLY	
AipaensisP6A5Q	ILKE*	T RDSFDLY	
MtnCADH	TILKN*	T RDSFDLY	ADH activity
Gglabra8888	ILKE*	T RDSFDLY	
Bbituminosa16676	ILKE*	T RDSFDLY	
GmncADH	ILKE*	T RDSFDLY	
SolyncADH	ILKE*	T RDSFDLY	
Graimondii06g129300.1	KLKE*	ATNDSFDLY	
Tcacao1EG01064111	KLKE*	TTNDSFDLY	
Graimondii09g126800.1	KLKE*	TTNDSFDLY	
Tcacao1EG02729011	ILVENT*	KGDSFDLY	
AtADH2	ILVENT*	KGDSFDLY	
Slycopersicum07g007590	DLVENT*	SSDSFDLY	
Slycopersicum09g011870	ILVENT*	NTASDSFDLY	
Fvesca17779	ILMENT*	TSDSFDLY	
Athaliana5g34930.1	ILVENT*	GDSDLY	
Fvesca30772	KLVENT*	TAGDSFDLY	
Graimondii10g059700.1	ILVENT*	TAGDSFDLY	
Tcacao1EG038186t1	ILVENT*	TAGDSFDLY	
Gmax17g13151	ILVENT*	TAGDSFDLY	
Gmax05g07870	SLVENT*	TAGDSFDLY	
Mtruncatula4g115980	SLVENT*	TAGDSFDLY	
Pvulgaris03g178200	SLVENT*	TGGDSFDLY	

Fig. 3B



C

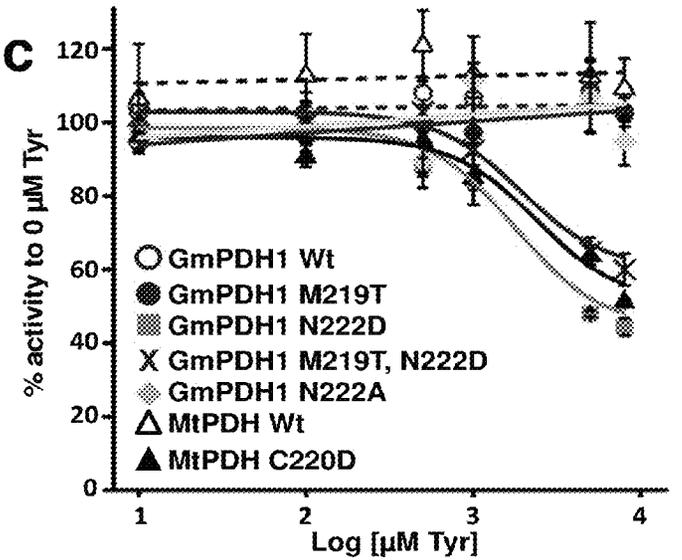


Fig. 3C

Fig. 4A

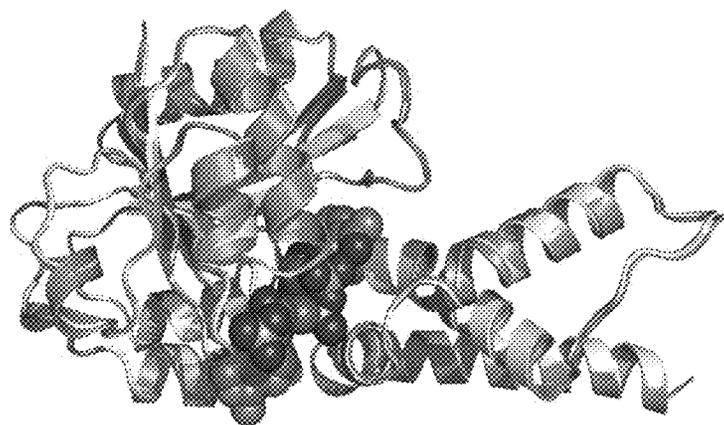


Fig. 4B

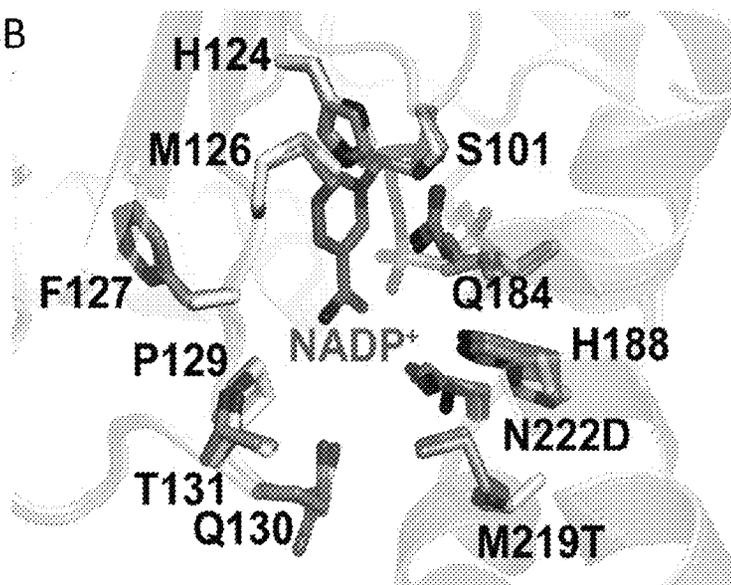


Fig. 4C

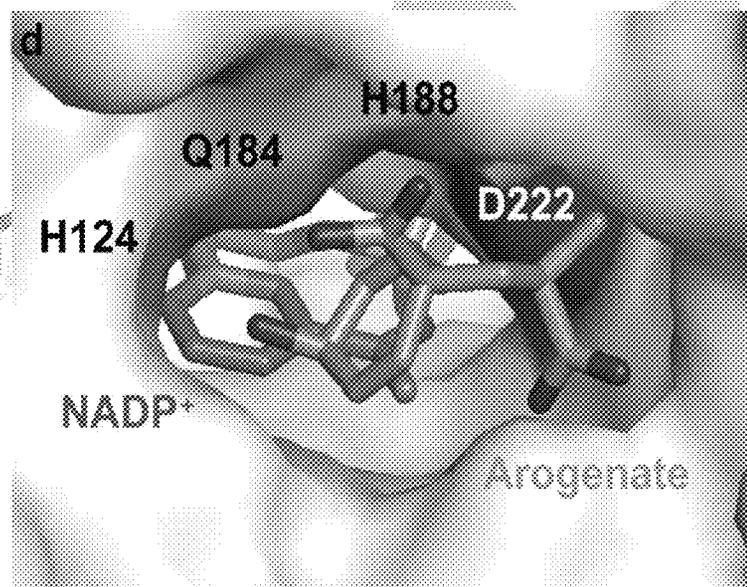
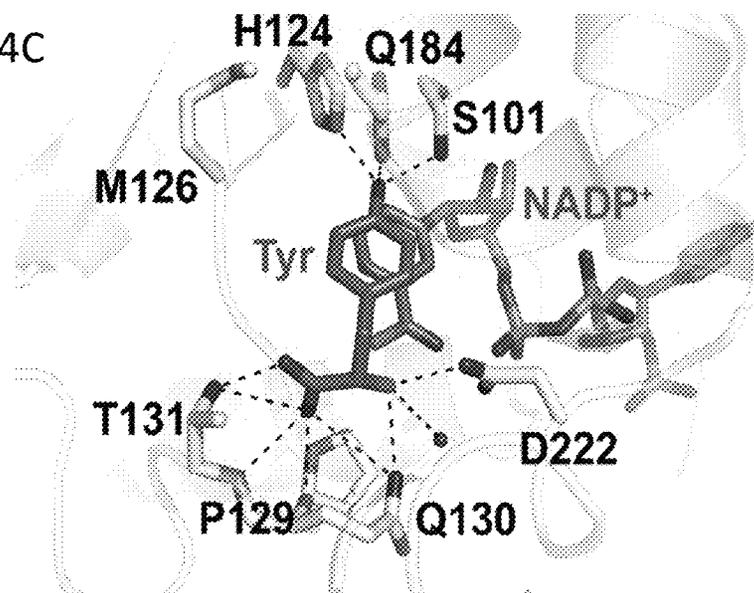


Fig. 4D

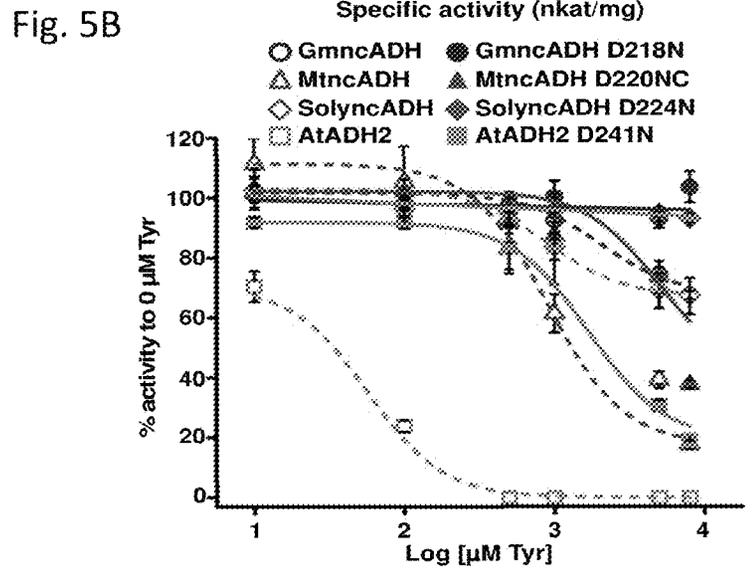
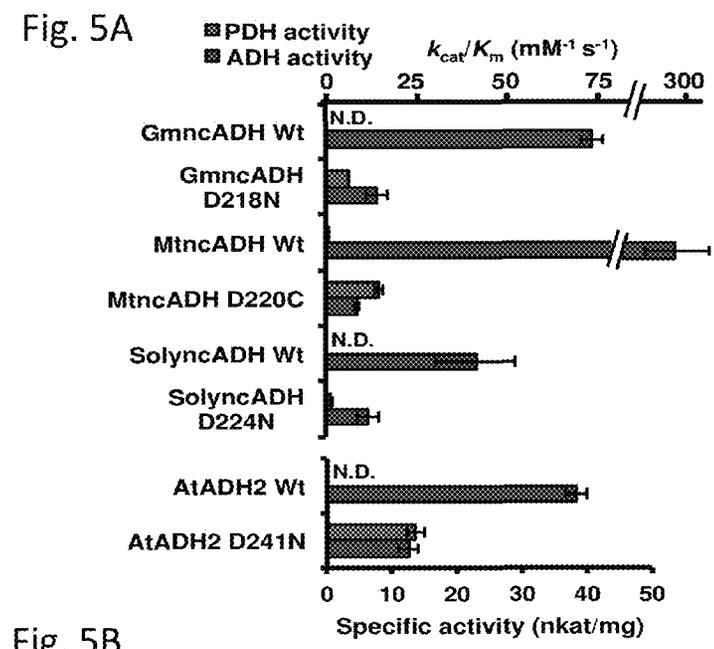


Fig. 6

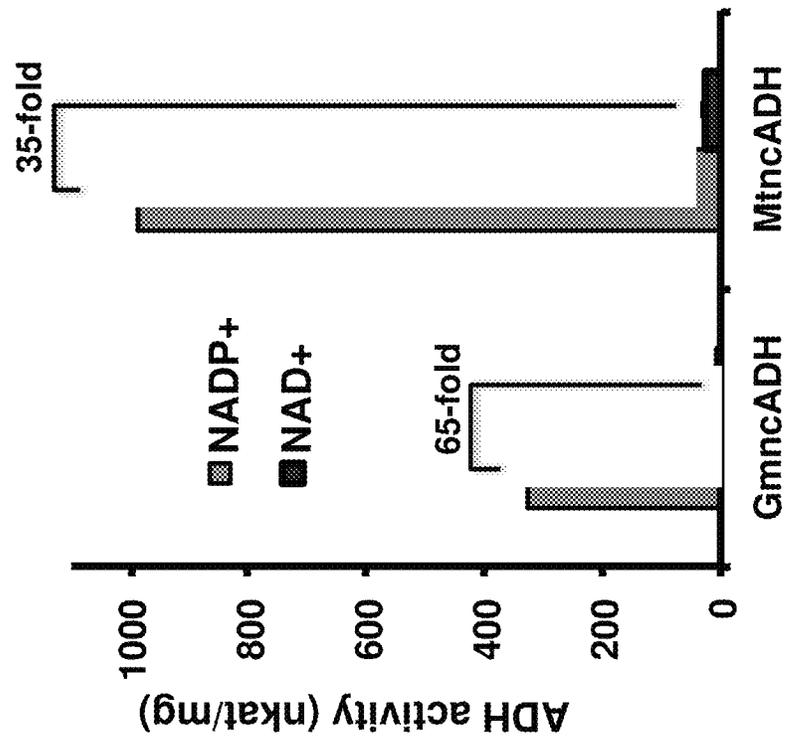


Fig. 7A

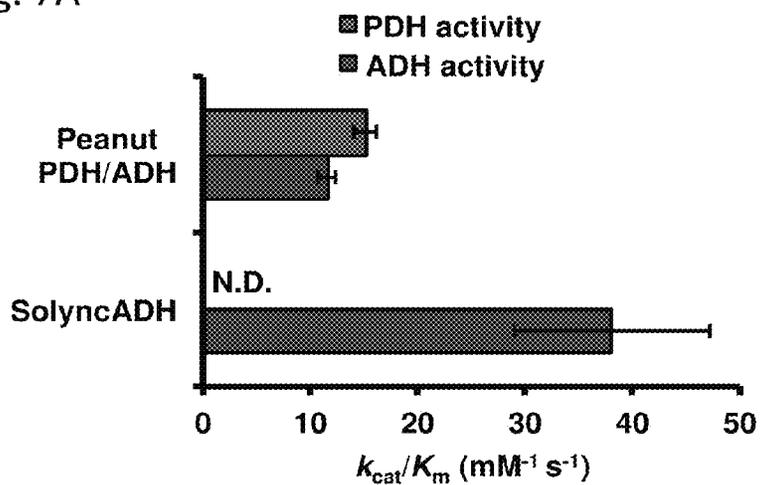


Fig. 7B

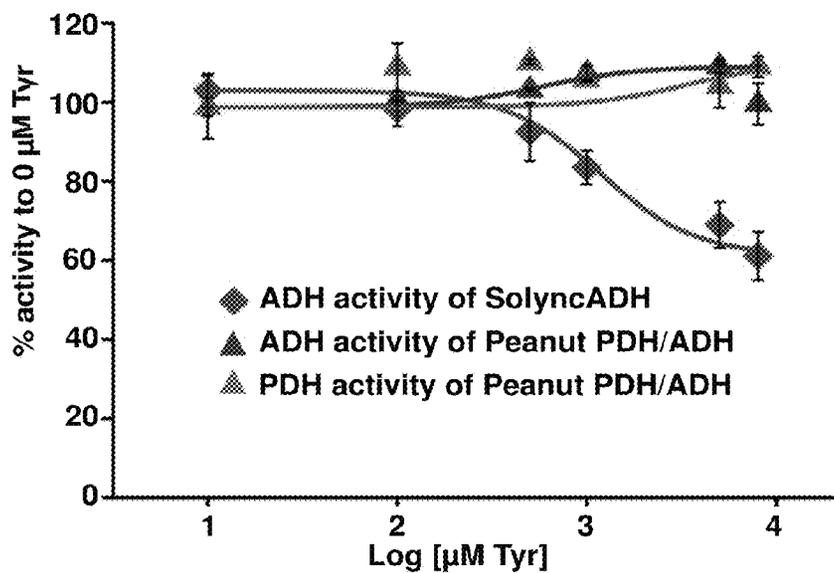


Fig. 8

GmPDH1	1	-----
Gm1g35760	1	-----
Bb68922	1	-----
Pv01g247200	1	-----
Gg9954	1	-----
Am5597	1	-----
MtPDH	1	-----
Peanut_PDH/ADH	1	-----
AdL1RYW	1	-----
AdP7853	1	-----
AiP6A5Q	1	-----
MtnoADH	1	-----
Gg8888	1	-----
Bb16676	1	-----
GmnoADH	1	-----
SolyncADH	1	-----
Gx06g129300.1	1	-----
Tc1EG040641t1	1	-----
Gx09g126800.1	1	-----
Tc1EG027290t1	1	-----MIKKNE
AtADH2	1	-----
Sl07g007590	1	-----
Sl09g011870	1	-----
Fv17779	126	LCTSIISTESVLSKSLPLQSLTHDTLFDVLDLVPVKEFSKALLLKMILPSNLAVVCTNPMFGVE--RQKHENWG
At5g34930.1	114	LCTSIIEIENILKTLFPQRLRRNTLFDVLSVKEFKAKTLLQYLPSDFDILCTHMPFGPQSVSSNHWNRG
Fv30772	139	LCTSIISAAAVLSSLPVQRRLRSTLFDVLSVKEFSKALFLMALPFYFDLICTHMPFGPE--SAKESWNE
Gx10g059700.1	136	LCTSIISTEQVLKTLPLQRLKRSSTLFDVLSVKEFKARNLLDALPLDFDIICSHPMFGPQ--SAKQSWKD
Tc1EG038186t1	141	LCTSIISTEQVLSKSLPLARLKRSTLFDVLSVKEFKAKTLLQYLPSDFDILCTHMPFGPQ--SAKQSWEN
Gm17g13151	138	LCSSIIISTQRVLLTFLQRLKRSTLFDVLSVKEFKAKTLLQYLPSDFDILCTHMPFGPQ--SAPRAWTG
Gm05g07870	135	LCSSIIISTQRVLLTFLQRLKRSTLFDVLSVKEFKAKTLLQYLPSDFDILCTHMPFGPQ--SAPRAWTG
Mt4g115980	134	LCTSIISAQVLLSLFPQRLKRSTLFDVLSVKEFKAKTLLQYLPSDFDILCTHMPFGPE--SCSSGWWG
Pv03g178200	139	LCSSIIISTQRVLLTFLQRLKRSTLFDVLSVKEFKAKTLLQYLPSDFDILCTHMPFGPQ--SAPRAWAG
GmPDH1	1	-----
Gm1g35760	1	-----
Bb68922	1	-----
Pv01g247200	1	-----
Gg9954	1	-----
Am5597	1	-----
MtPDH	1	-----
Peanut_PDH/ADH	1	-----
AdL1RYW	1	-----
AdP7853	1	-----
AiP6A5Q	1	-----
MtnoADH	1	-----
Gg8888	1	-----
Bb16676	1	-----
GmnoADH	1	-----
SolyncADH	1	-----
Gx06g129300.1	1	-----
Tc1EG040641t1	1	-----
Gx09g126800.1	1	-----
Tc1EG027290t1	7	DATFLPSVHLHLTKLLLVGSLKPLNLTQKKNVVKVVEPQPTTSBTKLSPQREGKINSPLKYSIPQLGEAN
AtADH2	1	-----
Sl07g007590	1	-----
Sl09g011870	1	-----NFS
Fv17779	194	SFLVYKVRIGSDS-VGISLCHDFLNI FEKESCRMVEMSCSVQDKYVAGSQFITHTVGRVLEKLTLESTP
At5g34930.1	184	LFVYKVRIGSER-LEVSRCESFLEIFVREGCRMVEMSVTDNDKFAAAGSQFITHTLGRLLGMLKLISTP
Fv30772	207	LHFVYKVRIGEG-SRAERCEKFLSIFEREGCRMVEMSCVDEHKFAAAGSQFITHTVGRVLEMLTVESTP
Gx10g059700.1	204	LFVYKVRIGNES-SRVQRCEQFLSIFEREGCRMVEMSCQEHDKFAAAGSQFMTHTVGRVLDMLGLESTP
Tc1EG038186t1	209	LFFVYKVRIGNES-SRVQRCEQFLSIFEREGCRMVEMSCQEHDKFAAAGSQFMTHTVGRVLDMLGLESTP
Gm17g13151	206	LPFVYKVRIGAHDDRRIARCEKFLSIFAREGCRMVEMSCADHDKFAAAGSQFITHTVGRVLEMLTVESTP
Gm05g07870	203	LPFVYKVRIGSDE-HRIARCEKFLSIFAREGCRMVEMSCADHDKFAAAGSQFITHTVGRVLEMLTVESTP
Mt4g115980	202	LPFVYKVRIGNE-TVSRCEKFLSIFAREGCRMVEMSCADHDKFAAAGSQFITHTVGRVLEMLTVESTP
Pv03g178200	207	LPFVYKVRIGDD-RRIERCEKFLSIFAREGCRMVEMSCADHDKFAAAGSQFITHTVGRVLEMLTVESTP

Fig. 8 cont.

GmPDH3	1	-----
Gm1g35760	1	-----
Bb68922	1	-----
Fv01g247200	1	-----
Gg9954	1	-----
Am5597	1	-----
MtPDR	1	-----
Peanut_PDR/ADH	1	-----
AdL1RYW	1	-----
AdP7PE3	1	-----
AlP6ASQ	1	-----
MtncADH	1	-----
Gg8888	1	-----
Bb16676	1	-----
GmncADH	1	-----
SclyncADH	1	-----
Gr06g129300.1	1	-----
Tc1EG040641t1	1	-----
Gr08g126800.1	1	-----
Tc1EG027290t1	1	-----
AtADH2	1	-----
Sl07g007590	1	-----
Sl09g011870	1	-----
Fv17779	1	-----
Ku5g34930.1	1	-----
Fv30772	1	-----
Gr10g059700.1	1	-----
Tc1EG038186t1	1	-----
Gm17g13151	1	-----
Gm05g07870	1	-----
Mt4g115985	1	-----
Fv03g178200	1	-----
GmPDH1	1	-----
Gm1g35760	1	-----
Bb68922	1	-----
Fv01g247200	1	-----
Gg9954	1	-----
Am5597	1	-----
MtPDR	1	-----
Peanut_PDR/ADH	1	-----
AdL1RYW	1	-----
AdP7PE3	1	-----
AlP6ASQ	1	-----
MtncADH	1	-----
Gg8888	1	-----
Bb16676	1	-----
GmncADH	1	-----
SclyncADH	1	-----
Gr06g129300.1	1	-----
Tc1EG040641t1	1	-----
Gr08g126800.1	1	-----
Tc1EG027290t1	1	-----
AtADH2	1	-----
Sl07g007590	1	-----
Sl09g011870	1	-----
Fv17779	56	LTEFHNVMKIAIIGFGNMQSLAKTFAHLGHIVLAHSRCEHENMAQDLGISYFHNQELCEQHPQVIL
At5g34930.1	47	---SNATPPLRIAIIGFGNYGQFLAETLISQGHILFAHSRSDHSSAAARLGVSYFTDLHDLCERHPQVIL
Fv30772	69	AAQFRSSQTLTIAIIGFGNYGQFLAKTLTQGHITVLAHSRSDHSSAAAKLGVSYFTDHLCEQHPEVIL
Gr10g059700.1	66	EDRAPPISLTKIAIIGFGNYGQFLANTLVSGGHIVLAHSRSDYSYVAKELGVSYFTDHLCEQHPEVIL
Tc1EG038186t1	71	RLAFSVNSTLKIAGIIGFGNYGQFLAKTLVAGGHISLLAYSRSRDSHSHIAQRLGVSYFTDHLCEQHPEVIL
Gm17g13151	68	AQQFHDQKIKIAIIGFGNFGQFLAQTIVRQGHITVLAHSRSDHSSAAQQLGVYFTFQHPDLCEHPEVIL
Gm05g07870	65	AQQFHDQKIKIAIIGFGNFGQFLAQTIVRQGHITVLAHSRSDHSSAAQQLGVYFTFQHPDLCEHPEVIL
Mt4g115980	64	ALQFHNQKIKIAIIGFGNFGQFLATTIVRQGHITVLAHSRSDYSVAQKIGVYFTFQHPDLCEHPEVIL
Fv03g178200	69	AQQFHDQKIKIAIIGFGNFGQFLAQTIVRQGHITVLAHSRSDYDAARKIGVYFTFQHPDLCEHPEVIL

Fig. 8 cont.

GmPDH1	1	-----
Gm1g35760	1	-----
Bb68822	1	-----
Pv01g247200	1	-----
Gg9954	1	-----
Am5597	1	-----
MtPDH	1	-----
Peanut_PDH/ADH	1	-----
AdLIRYW	1	-----
AdP7PS3	1	-----
AlP6A5Q	1	-----
MtncADH	1	-----
Gg8888	1	-----
Bb16676	1	-----
GmncADH	1	-----
SoLyncADH	1	-----
Gr06g129300.1	1	-----
Tc1EG040641t1	1	-----
Gr09g126800.1	1	-----
Tc1EG027290t1	77	-----MLIFSPFTIRGVNPPQLPFLSPSNSTSLPLSLTSLPTFSKSPFPSPHY
AtADH2	1	-----MLIHFSFAKELISPPNLRKHSPTFTISFFH-----SLRT
S107g007590	1	-----MLSFYPLGSKKFTFSNSNKKFSLNPTSTSTSRHHSVSSSSQVSHNKGKRLSL
S109g011870	4	-----LSAIQSNNYQSQSSSLLFNHHRHSTTISTPFPHHRLLFLPRAQNSDITATATKNNYVDLDDMLTRDKF
Fv17779	263	-----INTKGYETLLDLVDNLAGDSFDLYYGLFYNNKMAZEMPDRLDLAFWLRKQLFGLHLDIYRKLQFSSAK
At5g34930.1	253	-----INTKGYEALDLAENTCGDSFDLYYGLFYNNKNSLEVLRLDIAFAEALRKLFSRLNGVVRKQSPFGKAK
Fv30772	276	-----INTKGYETLLDLVNTAGDSFDLYYGLFYNNKALCOMLERLDEAFKALRAQLFGRLEHVVVRKQFPENAGK
Gr10g059700.1	273	-----INTNGYETLLDLVNTAGDSFDLYYGLFYNNKSALEMVERLDDAFDRLNKLFSRLHVVVRKQFPENAGK
Tc1EG038186t1	278	-----INTKGYETLLDLVNTAGDSFDLYYGLFYNNKSALEMVERLDDAFDRLNKLFSRLHVVVRKQFPENAGK
Gm17g13151	476	-----INTNGYELLLVNTAGDSFDLYYGLFYNNKNSLEMLERLDFAFEBLRKQLMARLHVVVRKQLSCHAKI
Gm09g07870	272	-----INTKGYEELLLVNTAGDSFDLYYGLFYNNKNSLEMLERLDFAFEBLRKQLMARLHVVVRKQLSCHAKI
Mt4g115980	271	-----INTKGYEELLLVNTAGDSFDLYYGLFYNNKNSLEMLERLDFAFEBLRKQLMARLHVVVRKQLSCHAKI
Pv03g178200	276	-----INTKGYEELLLVNTAGDSFDLYYGLFYNNKNSLEMLERLDFAFEBLRKQLMARLHVVVRKQLSCHAKI
GmPDH1	1	-----
Gm1g35760	1	-----
Bb68822	1	-----
Pv01g247200	1	-----
Gg9954	1	-----
Am5597	1	-----
MtPDH	1	-----
Peanut_PDH/ADH	1	-----
AdLIRYW	1	-----
AdP7PS3	1	-----
AlP6A5Q	1	-----
MtncADH	1	-----
Gg8888	1	-----
Bb16676	1	-----
GmncADH	1	-----
Gr06g129300.1	1	-----
Tc1EG040641t1	1	-----
SoLyncADH	1	-----
Gr09g126800.1	50	N-----
Tc1EG027290t1	147	Y-----
AtADH2	35	R-----
S107g007590	56	K-----
S109g011870	74	S-----
Fv17779	133	VBAEYVCH-----
At5g34930.1	323	KVHVFENC-----GEM-----D
Fv30772	346	VRFLEEN-----VN-----
Gr10g059700.1	343	VKSVRDS-----YQNGAASA-----
Tc1EG038186t1	348	GKSLGRK-----YQNSAGRA-----
Gm17g13151	346	GKVSLEFETVYQLLKKQWQSS-----AAAPLLS-----
Gm09g07870	342	GKVSLEFETVYVYQLLKKQWQNG-----SAAPVLS-----
Mt4g115980	341	VQNTLSEDSHRYVARRKNGM-----GSAIV-----
Pv03g178200	346	GKQSLEFETVYVYQLLKKQWQSSAPVLSLPSKESRSDVAKLIDKYVNTNDSQSDMNSKLMNALNGEAPVLS

Fig. 8 cont.

		Cofactor binding motif (GxGxxG)	NAD(P)+ specificity region
GaPDR1	1	MSSTSSG	LDLIGVGFQFQFLAKTMIKQGH
Ga11g13760	1	MTTMSSTSSG	LDKTIWGFQFQFLAKTMIKQGH
Bb64922	1	ZDTMSSTSSG	LDLIGVGFQFQFLAKTMIKQGH
Po61g247200	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
Gg9954	1	NSMSG	LDLIGVGFQFQFLAKTMIKQGH
Am5597	1	ASSG	LDLIGVGFQFQFLAKTMIKQGH
MLEPN	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
Peanut_PDH/ADR	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
AdL1R2W	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
AdP7P63	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
AlP6A5Q	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
MtncADR	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
Gg6888	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
Eh16676	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
GmncADR	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
SolyncADR	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
Gr06g129300.1	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
Tc1EG040641.t1	1	MSSSSG	LDLIGVGFQFQFLAKTMIKQGH
Gr06g126800.1	51	CTDAAGPPFYESHLANKYV	LDLIGVGFQFQFLAKTMIKQGH
Tc1EG027290.t1	148	SIDAAGPPFYESHLANKYV	LDLIGVGFQFQFLAKTMIKQGH
AtKAD2	36	AIDAAGPPFYESHLANKYV	LDLIGVGFQFQFLAKTMIKQGH
S107g007590	57	AIDAAGPPFYESHLANKYV	LDLIGVGFQFQFLAKTMIKQGH
S109g011870	75	KELSIENIENTELNPLCSHWK	LDLIGVGFQFQFLAKTMIKQGH
Fv17779	347	PSMGALRSHSQVLMQKARIGKRYDORSM	LDLIGVGFQFQFLAKTMIKQGH
At5g34830.1	335	ASLDMSSENVVYKYSQVSGSYDQSM	LDLIGVGFQFQFLAKTMIKQGH
Fv30772	156	MGALVRESVAVRFLKVVVLOSDDNSM	LDLIGVGFQFQFLAKTMIKQGH
Gx1G059700.1	159	FSSMALRSQVLIPIYEFKRIKESYVDSK	LDLIGVGFQFQFLAKTMIKQGH
Tc1EG039186.t1	364	SILNLSRSQVLIPIYEFKRIKESYVDSK	LDLIGVGFQFQFLAKTMIKQGH
Ga17g13151	176	LPSEDLRSQVLIPIYEFKRIKESYVDSK	LDLIGVGFQFQFLAKTMIKQGH
Gm05g07870	171	LPSEDLRSQVLIPIYEFKRIKESYVDSK	LDLIGVGFQFQFLAKTMIKQGH
Nt4g115980	364	LSSKNGRSADARLGGYTRNGSDQDDTK	LDLIGVGFQFQFLAKTMIKQGH
Pv03g178200	416	HVSKELAFDVTZLNKYPHISSD	LDLIGVGFQFQFLAKTMIKQGH

		Ser101	
GaPDR1	50	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Ga11g13760	53	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Bb64922	51	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Po61g247200	47	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Gg9954	46	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Am5597	48	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
MLEPN	49	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Peanut_PDH/ADR	49	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
AdL1R2W	49	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
AdP7P63	50	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
AlP6A5Q	47	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
MtncADR	45	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Gg6888	44	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Eh16676	46	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
GmncADR	52	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
SolyncADR	52	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Gr06g129300.1	52	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Tc1EG040641.t1	53	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Gr06g126800.1	113	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Tc1EG027290.t1	211	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
AtKAD2	99	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
S107g007590	120	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
S109g011870	138	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Fv17779	417	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
At5g34830.1	405	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Fv30772	426	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Gx1G059700.1	429	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Tc1EG039186.t1	434	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Ga17g13151	446	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Gm05g07870	441	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Nt4g115980	434	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH
Pv03g178200	486	SHVFFEDSAG	LDLIGVGFQFQFLAKTMIKQGH



Fig. 8 cont.

	Asn222	
GmPDI1	221	.....
Gm11g35760	224	.....
Bb68922	224	.....
Pv01g247200	222	.....
Gg9954	218	.....
An5597	217	.....
MtPDIH	219	.....
Peanut_PDIH/ADR	220	.....
AdL1RXW	232	.....
AdP7PE3	222	.....
AlP6ASQ	222	.....
NtnCADH	219	.....
Gg8888	216	.....
Bb16676	215	.....
GmncADH	217	.....
SolyacADH	223	.....
Gr09g129300.1	223	.....
Tc1EG040641t1	224	.....
Gr09g126800.1	288	.....
Tc1EG027290t1	396	.....
AtADH2	274	.....
Sl07g007590	295	.....
Sl09g011870	314	.....
Pv17779	587	.....
At5g34930.1	580	.....
Pv30772	601	.....
Gr10g059700.1	604	.....
Tc1EG038186t1	609	.....
Gm17g13151	621	.....
Gm05g07870	616	.....
Mt4g115860	609	.....
Pv03g178200	661	.....
GmPDI1	.....	.....
Gm11g35760	.....	.....
Bb68922	.....	.....
Pv01g247200	.....	.....
Gg9954	.....	.....
An5597	.....	.....
MtPDIH	.....	.....
Peanut_PDIH/ADR	.....	.....
AdL1RXW	.....	.....
AdP7PE3	.....	.....
AlP6ASQ	.....	.....
NtnCADH	.....	.....
Gg8888	.....	.....
Bb16676	.....	.....
GmncADH	.....	.....
SolyacADH	.....	.....
Gr09g129300.1	.....	.....
Tc1EG040641t1	.....	.....
Gr09g126800.1	346	.....
Tc1EG027290t1	443	.....
AtADH2	332	.....
Sl07g007590	353	.....
Sl09g011870	372	.....
Pv17779	637	.....
At5g34930.1	629	.....
Pv30772	652	.....
Gr10g059700.1	653	.....
Tc1EG038186t1	679	.....
Gm17g13151	671	.....
Gm05g07870	666	.....
Mt4g115860	654	.....
Pv03g178200	712	.....

Fig. 9A <sup>a</sup>

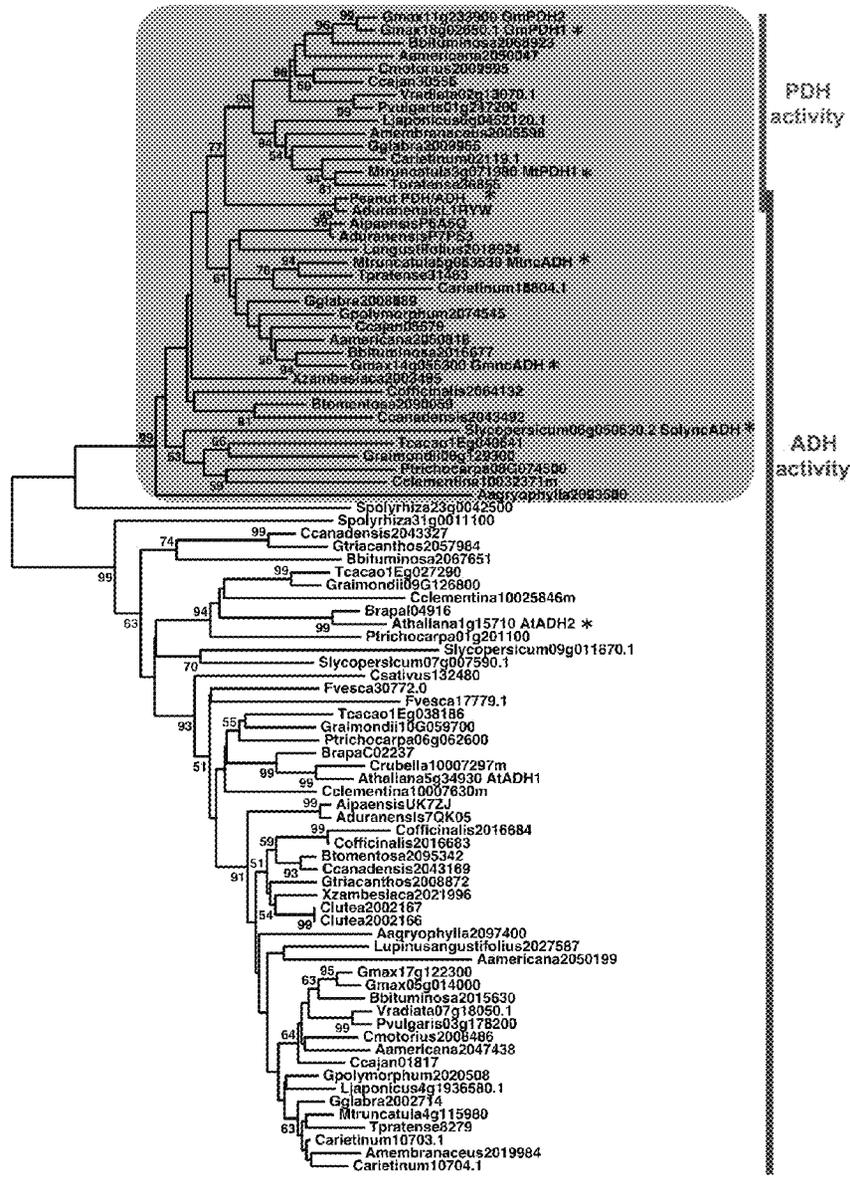


Fig. 9B

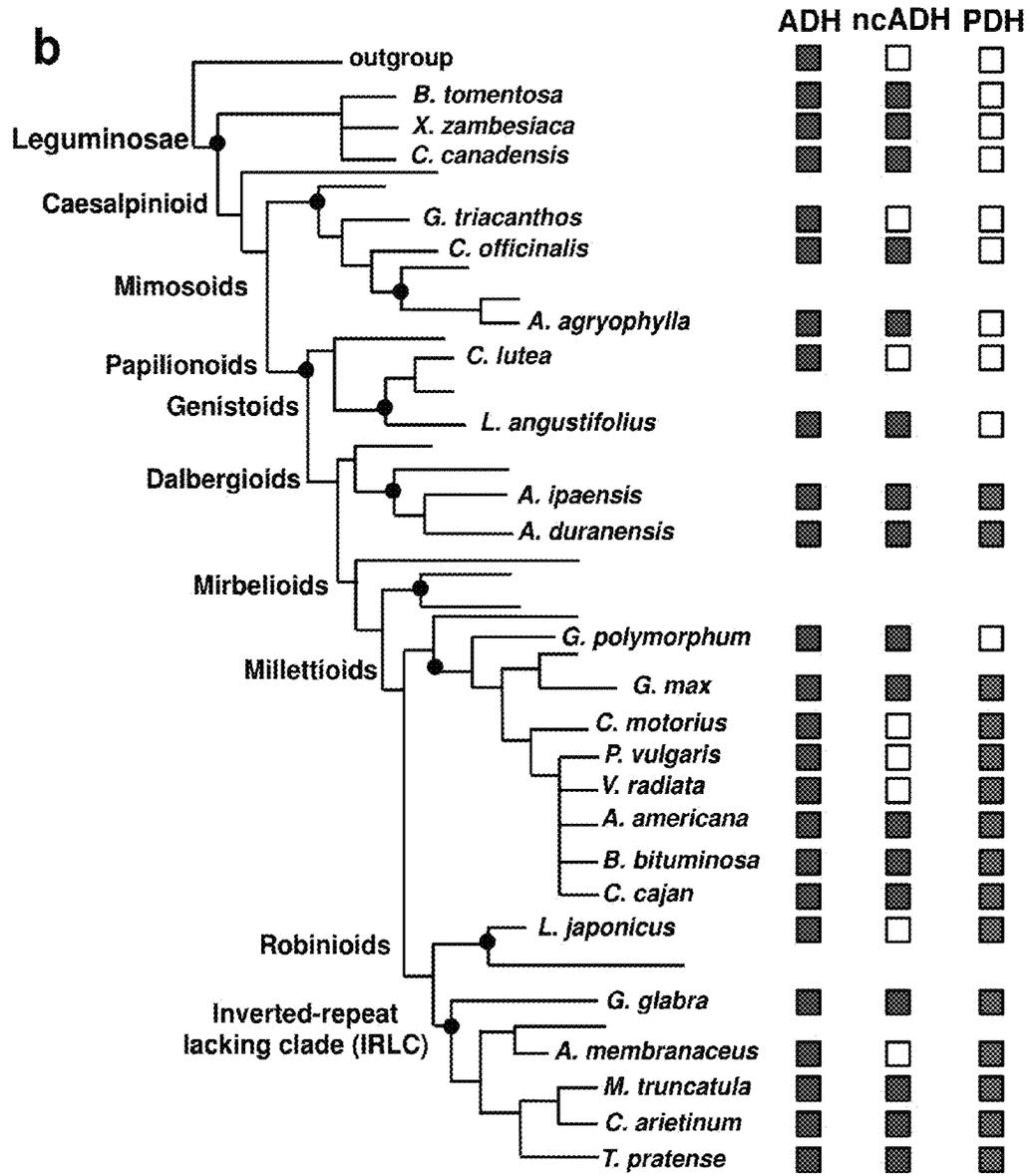
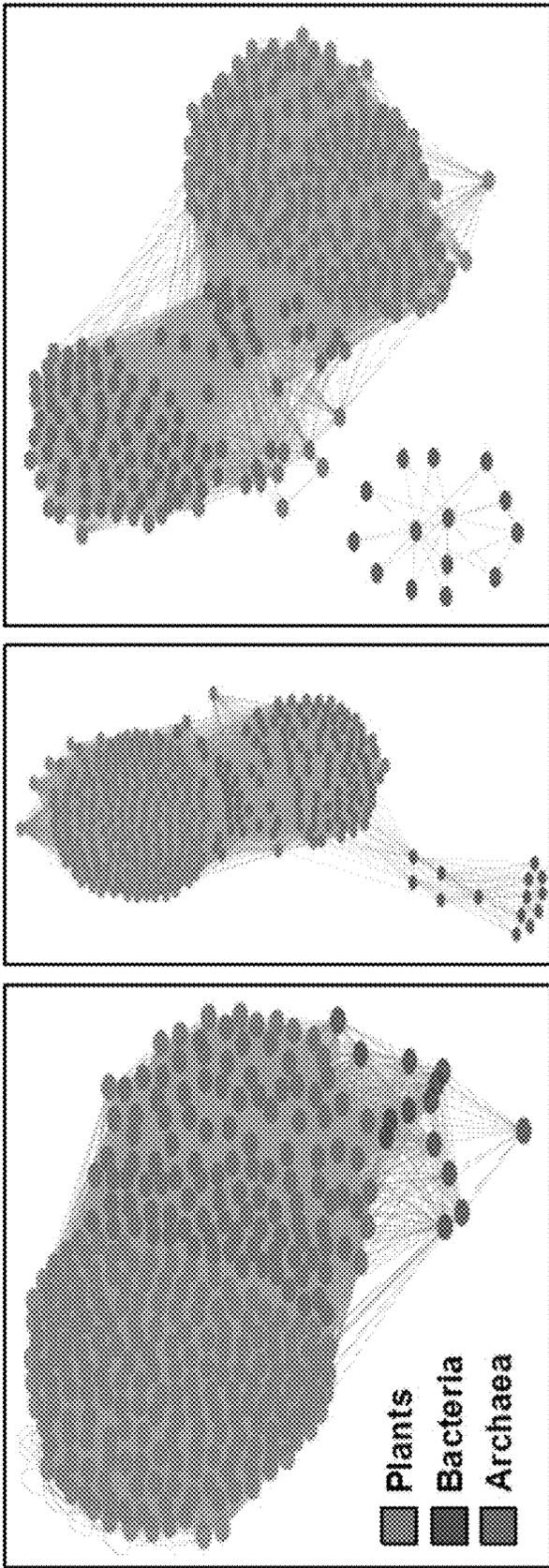


Fig. 10A



a

Fig. 10B

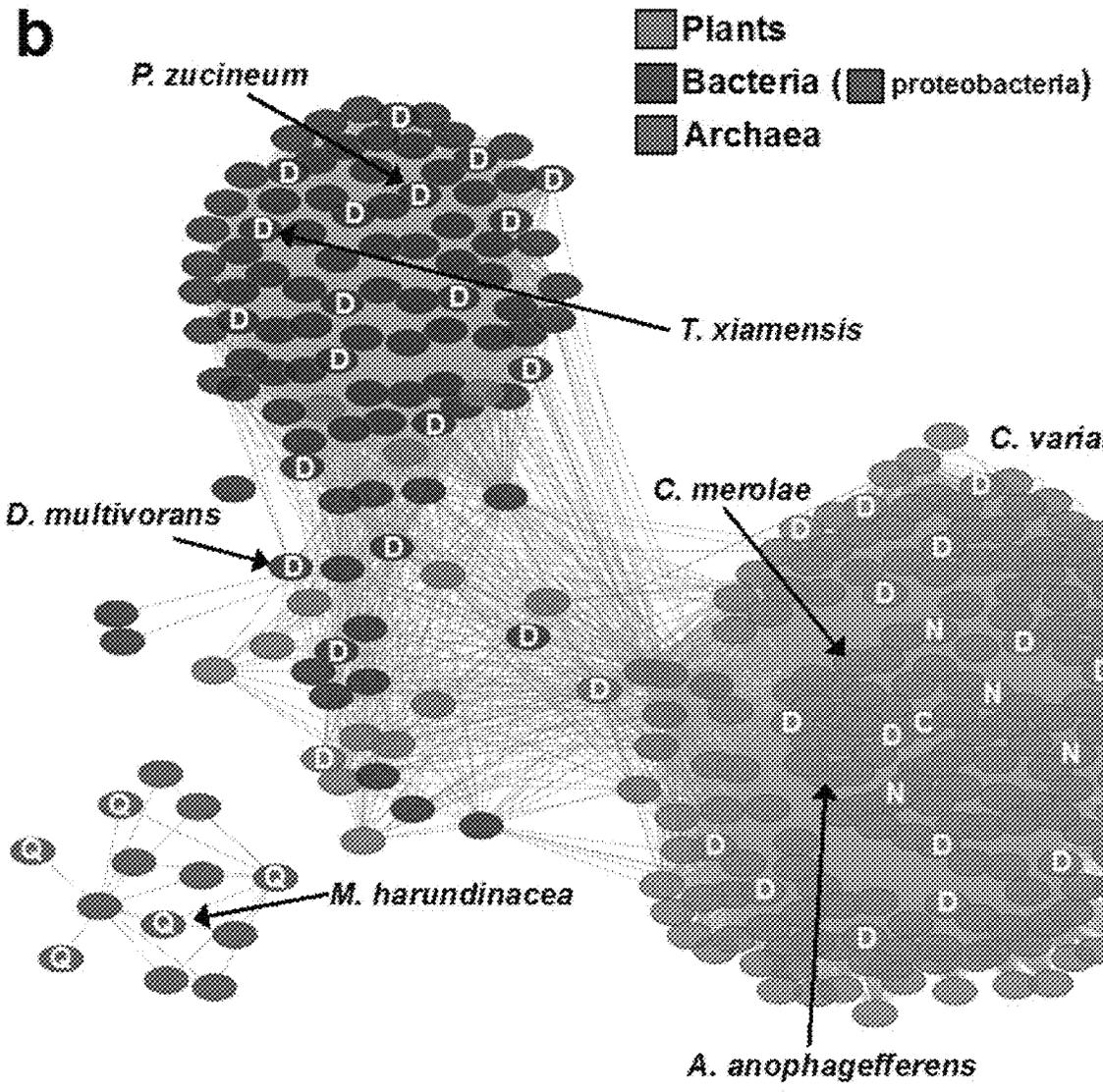


Fig. 10C

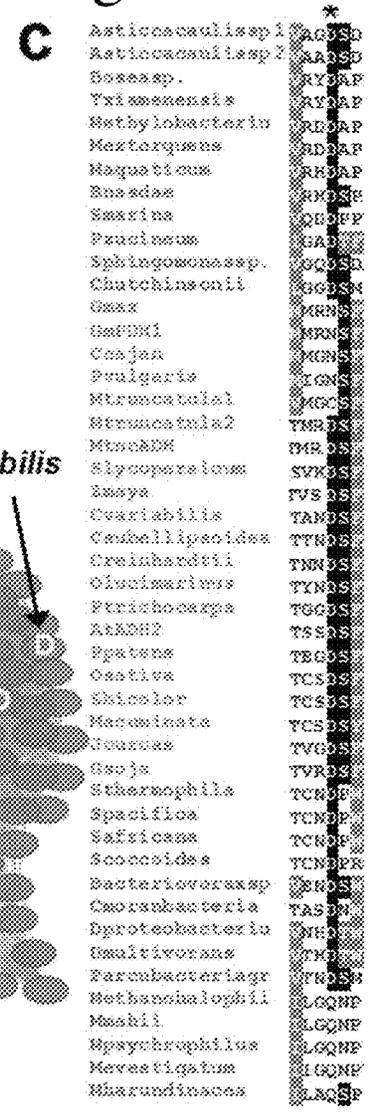


Fig. 11A

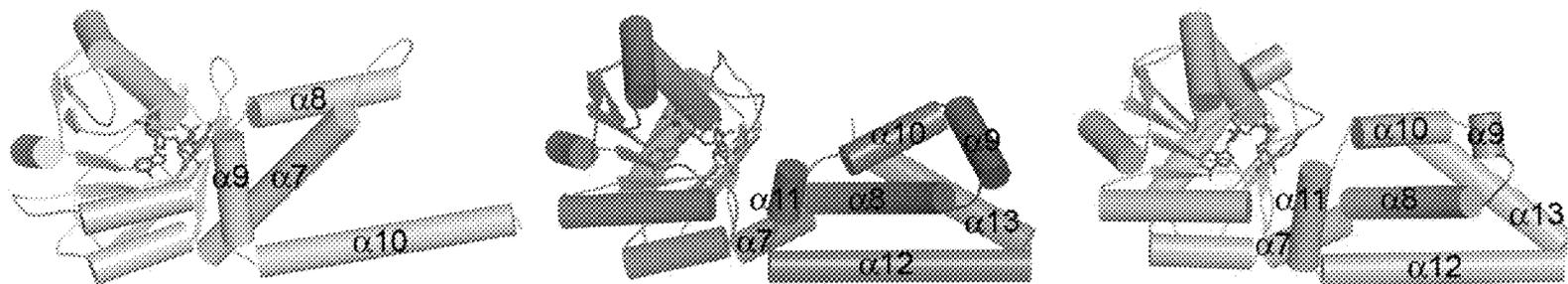


Fig. 11B

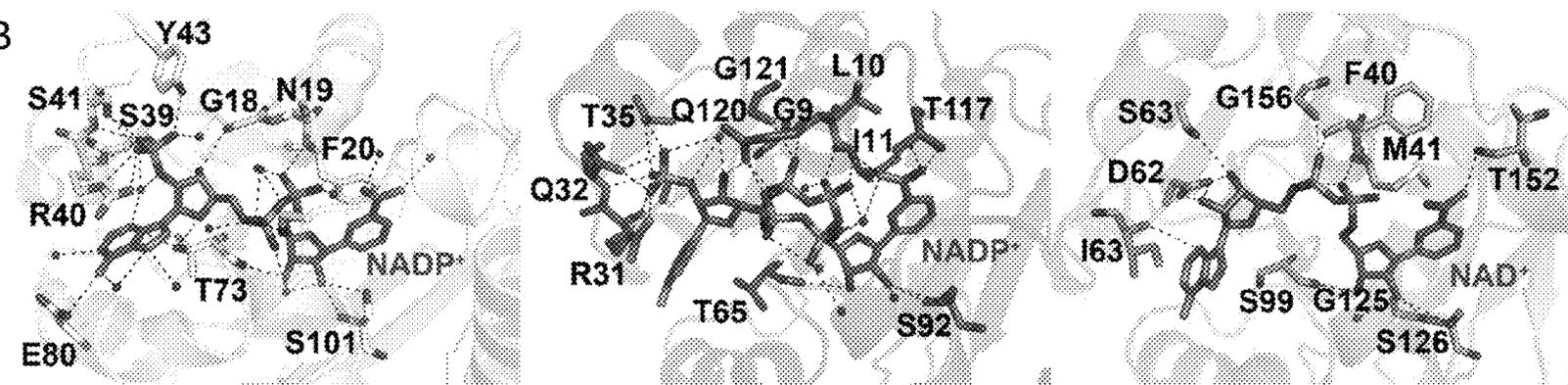
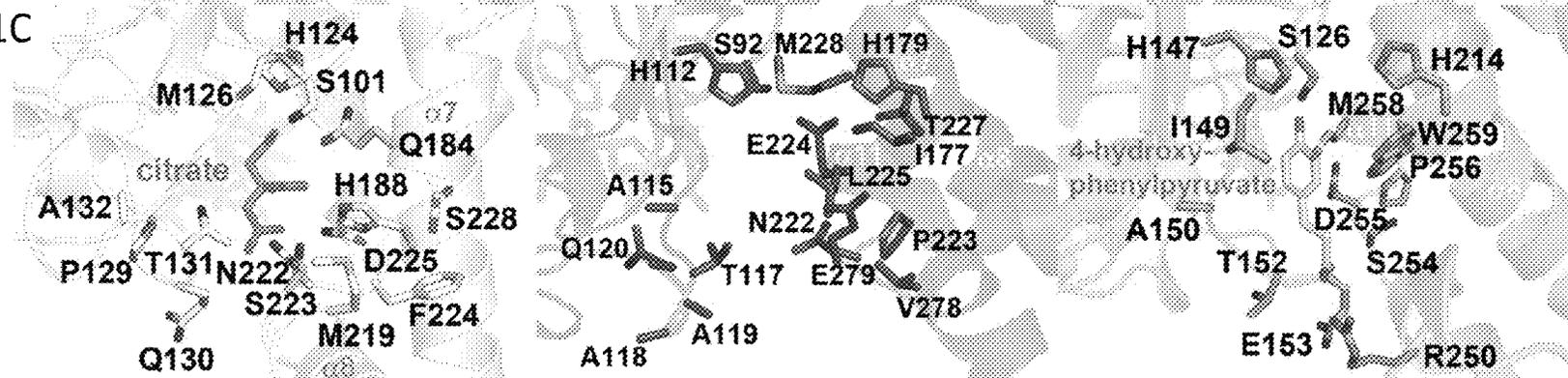


Fig. 11C





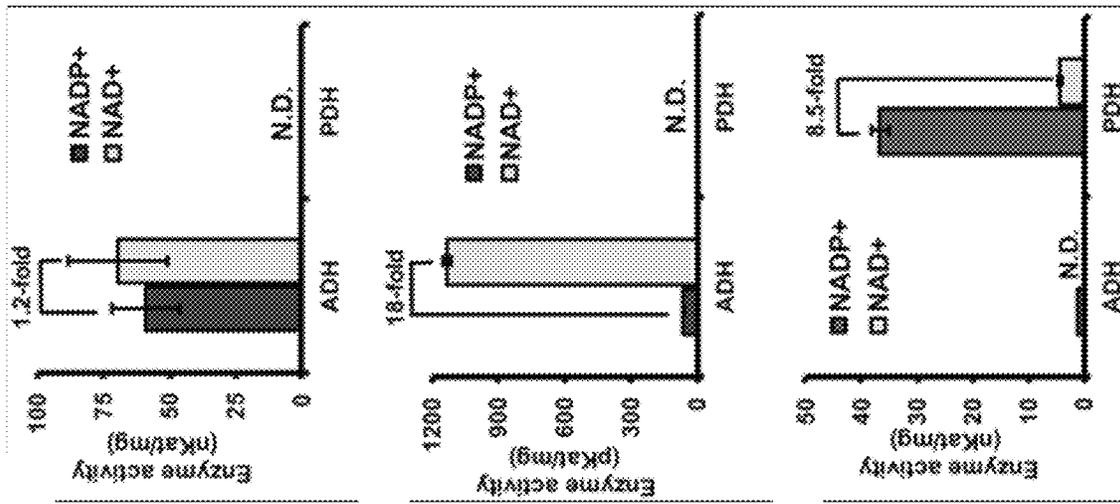


Fig. 13A

Fig. 13B

Fig. 13C

Fig. 14A

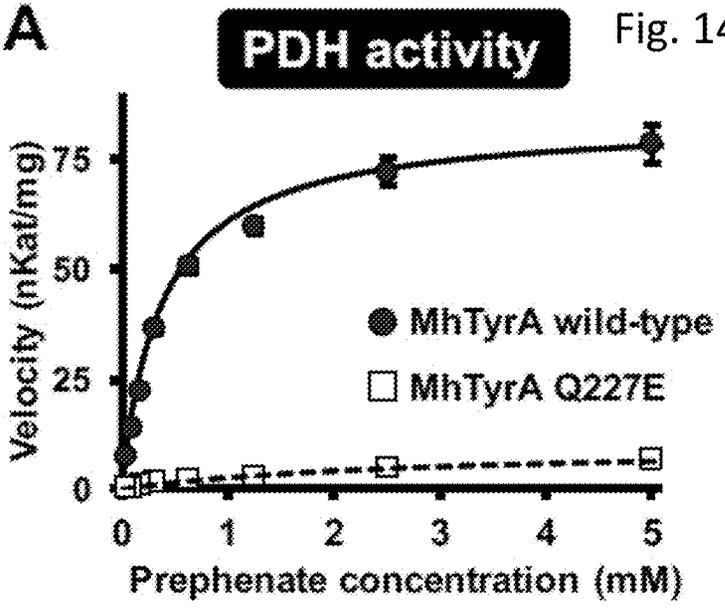


Fig. 14B

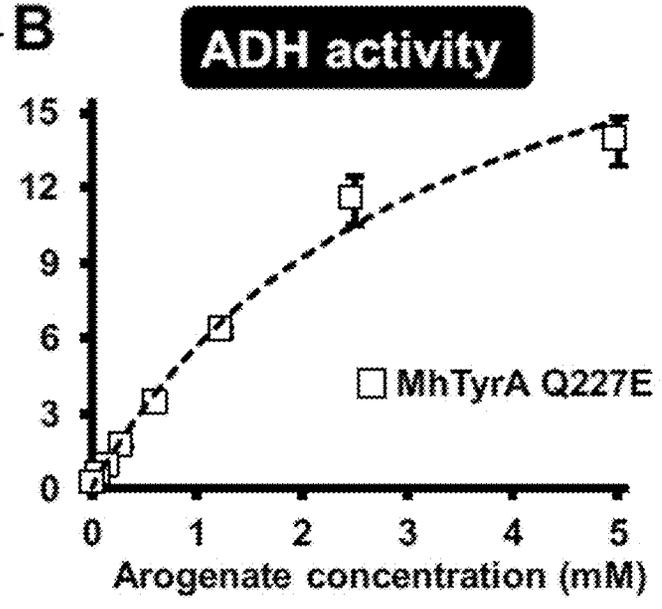


Fig. 15

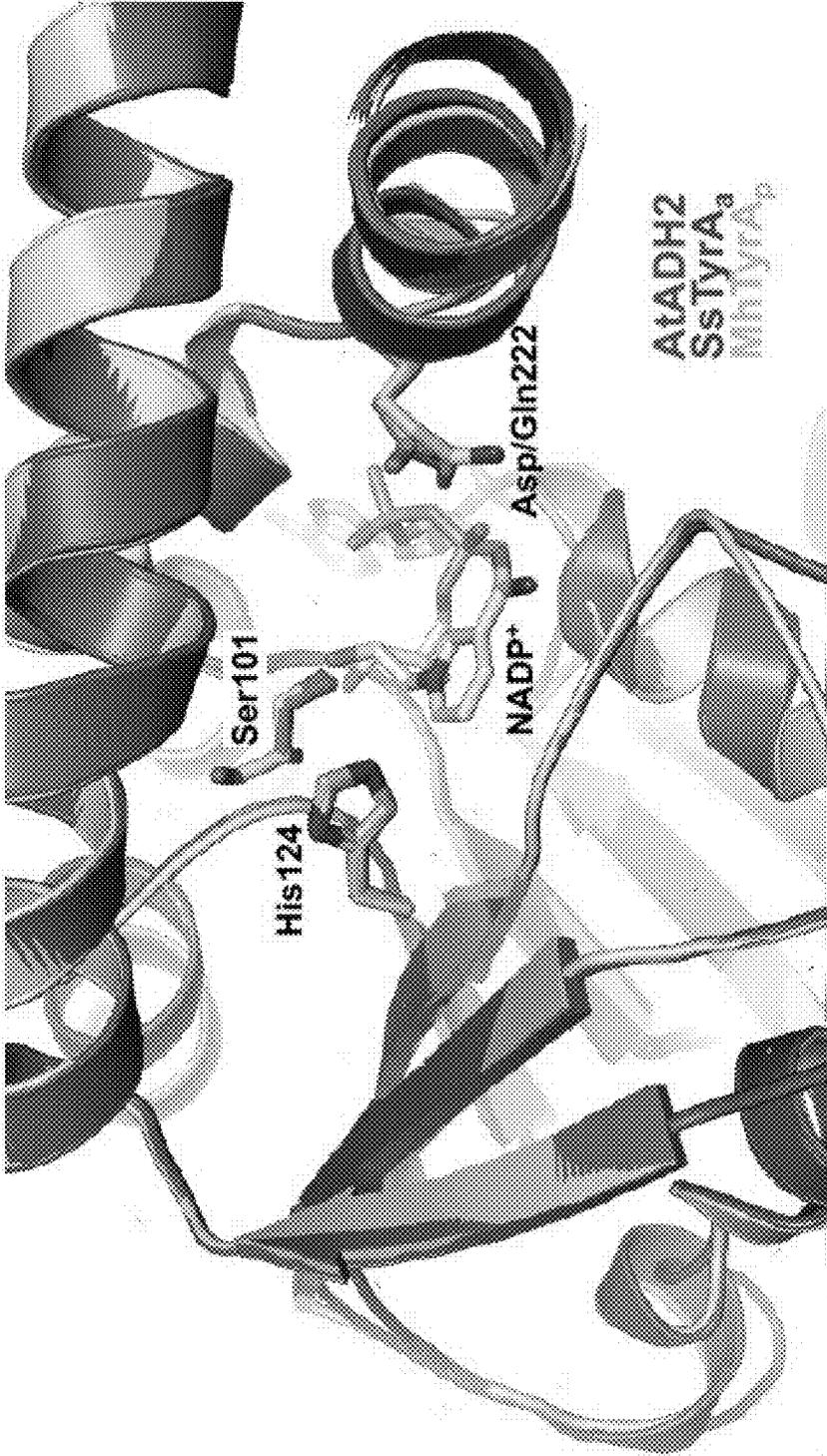


Fig. 16

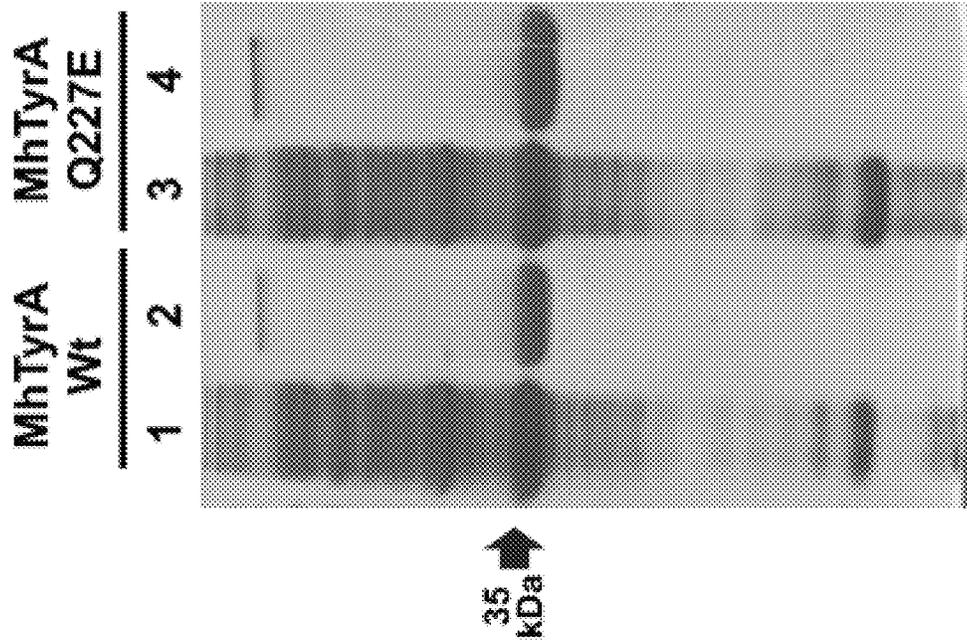


Fig. 17

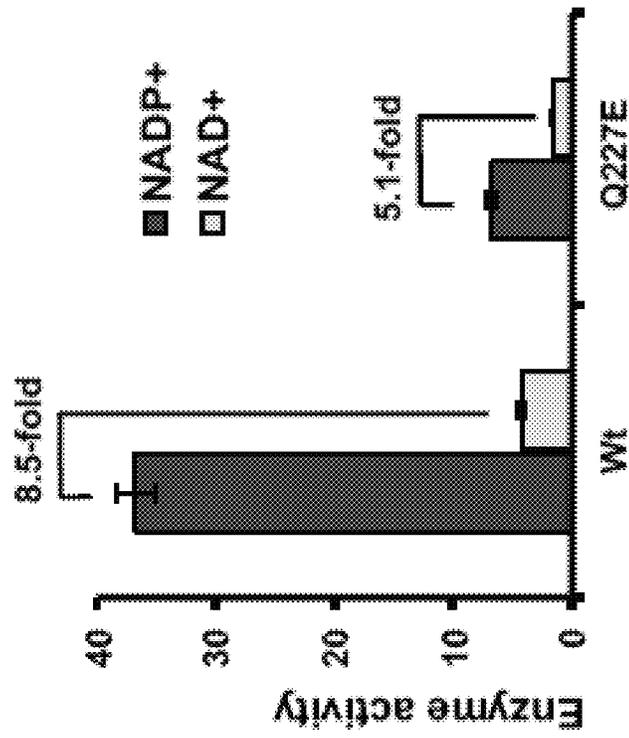
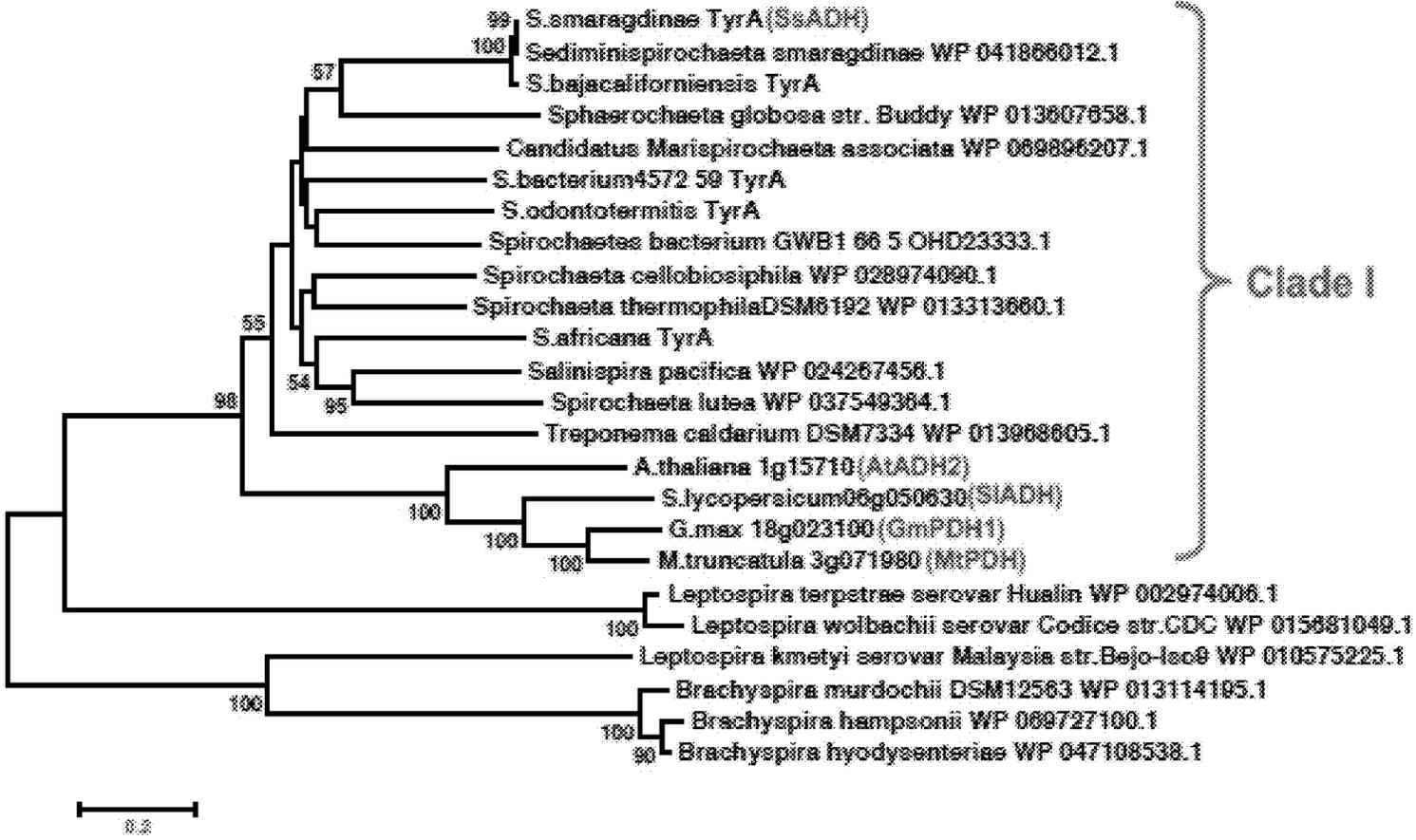


Fig. 18



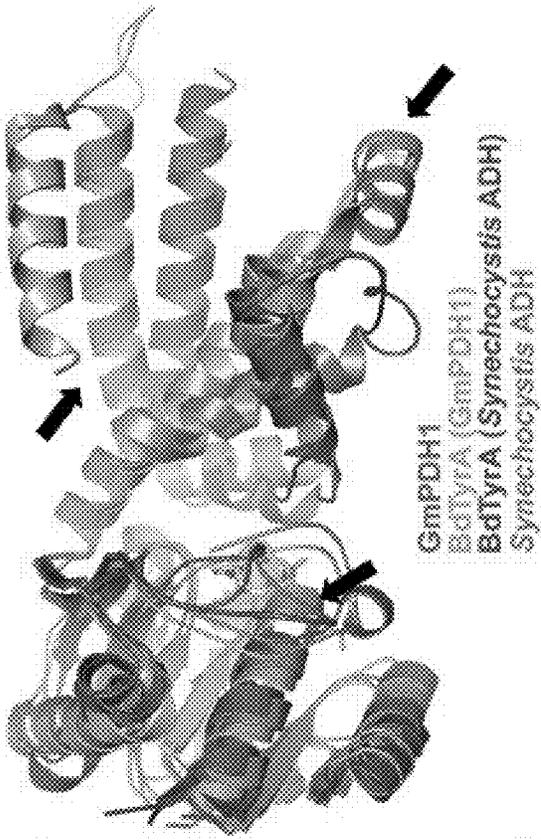


Fig. 19A

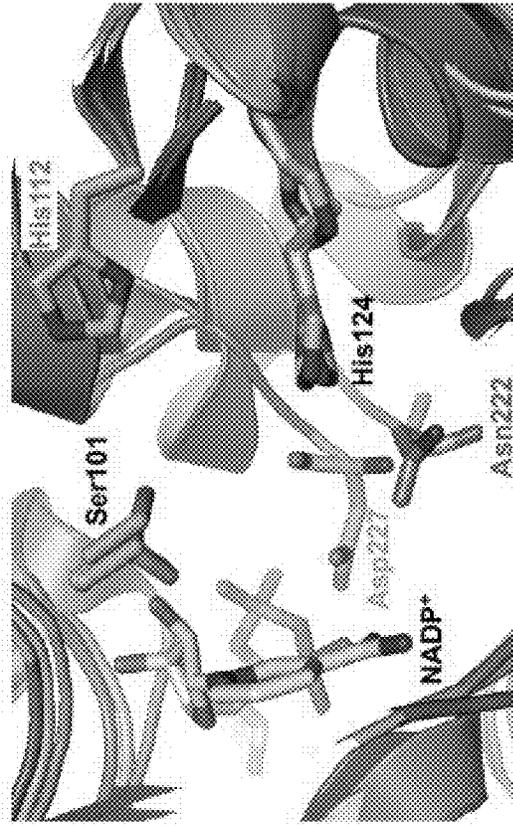


Fig. 19B

**ENGINEERED PREPHENATE  
DEHYDROGENASES AND AROGENATE  
DEHYDROGENASES AND METHODS OF  
USING THE SAME**

CROSS-REFERENCE TO RELATED PATENT  
APPLICATIONS

The present application claims the benefit of priority to U.S. Provisional Patent Application No. 62/451,124, filed on Jan. 27, 2017, the content of which is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY  
SPONSORED RESEARCH

This invention was made with United States government support awarded by the National Science Foundation grant number IOS-1354971. The United States has certain rights in this invention.

SEQUENCE LISTING

This application is being filed electronically via EFS-Web and includes an electronically submitted Sequence Listing in .txt format. The .txt file contains a sequence listing entitled "2018-01-29-\_5671-00080\_ST25.txt" created on Jan. 29, 2018 and is 668,439 bytes in size. The Sequence Listing contained in this .txt file is part of the specification and is hereby incorporated by reference herein in its entirety.

INTRODUCTION

L-Tyrosine (Tyr) is an essential aromatic amino acid required for protein synthesis in all organisms but, synthesized de novo only in plants and microorganisms. The Neurotransmitters such as catecholamines in metazoans are derived from Tyr, which must be obtained from their diet, as they cannot synthesize Tyr de novo<sup>8</sup>. In plants, Tyr serves as the precursor to numerous specialized metabolites crucial for both plant and human health, such as antioxidants vitamin E, the photosynthetic electron carrier plastoquinone, betalain pigments, and defense compounds, including dhurrin, rosmarinic acid, and isoquinoline alkaloids (e.g. morphine)<sup>9-14</sup>. The major plant cell wall component lignin can also be synthesized from Tyr in grasses<sup>15</sup>.

Tyr is synthesized from prephenate, a shikimate pathway product, by two reactions, an oxidative decarboxylation and a transamination. The TyrA enzymes catalyze the oxidative decarboxylation step and are the key regulatory enzymes of Tyr biosynthesis, as they are usually inhibited by Tyr and compete for substrates that are also used in L-phenylalanine biosynthesis. In many microbes an NAD(H)-dependent prephenate dehydrogenase/TyrA (PDH/TyrA<sub>p</sub>; EC 1.3.1.13) converts prephenate into 4-hydroxyphenylpyruvate (HPP) followed by transamination to Tyr by Tyr aminotransferase (TAT). In plants, these two reactions occur in the reverse order, with prephenate first being transaminated to aroenate by prephenate aminotransferase (PPA-AT), followed by oxidative decarboxylation to Tyr by an NADP(H)-dependent aroenate dehydrogenase/TyrA (ADH/TyrA<sub>a</sub>; EC 1.3.1.78)<sup>19-24</sup>. Some exceptions to these "textbook" models are found in nature including microbes that use ADH to synthesize Tyr<sup>25,26</sup> and plants such as legumes having PDH activity<sup>5,27,28</sup>. Also, some microbial TyrAs prefer NADP(H) cofactor<sup>18,9</sup>. Thus, variations exist in the TyrA enzymes in

diverse organisms, yet the molecular basis underlying TyrA substrate specificity and the alternative Tyr pathways is currently unknown.

Comparison of microbial TyrA sequences identified an aspartate residue downstream of the NAD(P)(H) binding motif that was later shown to confer cofactor specificity of TyrA<sup>16,30</sup>. Site-directed mutagenesis of *Escherichia coli* PDH and structural analysis of *Aquifex aeolicus* PDH identified an active site histidine, which interacts with substrate C4-hydroxyl and is critical for catalysis in each PDH. The same studies also showed that an active site arginine is necessary for substrate binding, but not for substrate specificity<sup>31-34</sup>. Besides their varied substrate and cofactor specificities, TyrA enzymes also exhibit different regulatory properties. Mutation of another active site histidine, which is present in the *E. coli* and *A. aeolicus* PDHs but absent in Tyr-insensitive *Synechocystis* ADH, relieved Tyr inhibition but simultaneously reduced PDH activity<sup>34</sup>. Random mutagenesis of the *E. coli* enzyme identified additional residues that relaxed Tyr inhibition; however, PDH activity was also reduced in these mutants<sup>35</sup>. Sequence and structural comparisons of divergent TyrA homologs, however, have been unable to identify specific determinants of Tyr-sensitivity and substrate specificity<sup>16,29,30,33,34</sup>.

Understanding the specific determinants of Tyr-sensitivity and substrate specificity in ADH or PDH enzymes would allow one to engineer new ADH or PDH polypeptides with unique properties that would be useful in producing important commercial products derived from the Tyr pathway. For example, betalains, important pharmaceuticals such as L-dihydroxyphenylalanine (L-DOPA), and benzylisoquinoline alkaloids such as morphine are synthesized from Tyr. Betalains are used as a natural food dye (E162) and have anticancer and antidiabetic properties. Consequently, there is a need in the art for new ADH or PDH polypeptides that may be used to enhance the production of Tyr in cells, and thus the yield of Tyr-derived plant natural products important for human health and nutrition.

BRIEF DESCRIPTION OF DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIGS. 1A-1D show the Tyr biosynthesis pathways in plants and identification and characterization of noncanonical ADHs from legumes. FIG. 1A shows two Tyr biosynthetic routes from prephenate. The PDH (blue, left) pathway is present in most microbes and legumes, whereas the ADH (red, right) pathway is ubiquitous in plants. Dashed line represents feedback inhibition by Tyr. FIG. 1B shows a phylogenetic analysis of TyrA homologs from various eudicot lineages identified a clade of ADH/PDH homologs (noncanonical, gray box) distinct from previously characterized plant ADH (canonical). Plant PDHs form a subgroup in the noncanonical clade. FIG. 1C is a graph showing PDH (blue; top) and ADH (red; bottom) activity of PDH, and noncanonical ADHs with NADP<sup>+</sup> cofactor from 4 plants. Catalytic efficiency ( $k_{cat}/K_m$ ) is expressed as  $\text{mM}^{-1} \text{s}^{-1} \pm \text{SEM}$  of  $n \geq 3$ . N.D., below detection limit. FIG. 1D is a graph showing the effect of Tyr on plant ADH and PDHs. Data are shown as IC<sub>50</sub> plots with enzymatic activity determined at increasing amounts of L-Tyr (0-8 mM). Activity was normalized to an assay with no L-Tyr and expressed as percent activity of  $n=3 \pm \text{SEM}$ . PPA-AT, Prephenate aminotransferase, TAT, Tyrosine aminotransferase.

FIGS. 2A-2D show the X-ray crystal structure of GmPDH1. FIG. 2A is a ribbon diagram showing the monomeric units (colored gold (left) and white (right), respectively) of the homodimer. NADP<sup>+</sup> (green) and citrate (purple) are depicted as space-filling models. The N- and C-terminal domains are also indicated. FIG. 2B is an electron density map for NADP<sup>+</sup>. The 2F<sub>o</sub>-F<sub>c</sub> omit map (1.5 σ) for the ligand is shown. FIG. 2C shows the nicotinamide cofactor binding pocket of GmPDH1. Residues surrounding the bound NADP<sup>+</sup> (green) and water molecules (red spheres) are shown. Ligand interactions are indicated by dotted lines. FIG. 2D shows the active site residues in GmPDH1 in contact with citrate (purple) to identify the proposed prephenate binding site.

FIGS. 3A-3C show the identification of Asn222 as a determinant of PDH activity and Tyr sensitivity. FIG. 3A shows a trimmed amino acid alignment corresponding to the phylogeny in FIG. 1B highlighting residues Met219 and Asn222 (number based on GmPDH1; See SEQ ID NOS: 169-200). FIG. 3B is a graph showing PDH (blue; top bar) and ADH (red; bottom bar) activity of GmPDH1, MtPDH and corresponding site-directed mutants. Bars represent average catalytic efficiency ( $k_{cat}/K_m$ ) in  $\text{mM}^{-1} \text{s}^{-1} \pm \text{SEM}$  of  $n=3$  replications. N.D., below detection limit. FIG. 3C is a graph showing the effect of Tyr on PDH activity of wild-type and mutant GmPDH1 and MtPDH. Data are shown as IC<sub>50</sub> plots with enzymatic activity determined at increasing concentrations of L-Tyr (0-8 mM). Activity was normalized to an assay with no L-Tyr and expressed as percent activity of  $n=3 \pm \text{SEM}$ . Open symbols correspond to wild-type enzymes, with dashed lines. Mutant enzymes have filled symbols with solid lines.

FIGS. 4A-4D show the crystal structures of GmPDH1 N222D and M219T/N222D to reveal Tyr binding interactions. FIG. 4A is a set of ribbon diagrams showing the overlay of GmPDH1 (blue), GmPDH1 N222D (rose), and GmPDH1 M219T/N222D (white) with NADP<sup>+</sup> (green) shown as a space-filling model. FIG. 4B is an active site overlay of wild-type and mutant GmPDH1 which shows a conserved architecture. Coloring of side-chains is the same as for panel A. FIG. 4C shows active site residues in GmPDH1 M219T/N222D in contact with Tyr (purple). FIG. 4D shows molecular docking of argenate (rose) into the active site of GmPDH1 M219T/N222D. The surface of the active site pocket is shown with the surface corresponding to Asp222 colored red.

FIGS. 5A-5B show that Asn222 confers PDH activity to divergent plant ADHs while simultaneously introducing Tyr sensitivity. FIG. 5A is a graph showing ADH activity from wild-type ADH enzymes and their mutants that remove Asp at the corresponding 222 position. Bars represent average catalytic efficiency ( $k_{cat}/K_m$ ) in  $\text{mM}^{-1} \text{s}^{-1} \pm \text{SEM}$  for  $n=3$ . Activity from AtADH2 is shown as specific activity (nkat/mg  $\pm \text{SEM}$  for  $n=3$ ) as kinetics were unable to be determined. N.D., below detection limit. FIG. 5B shows IC<sub>50</sub> plots analyzing Tyr sensitivity of ADH activity from wild-type and mutated ADHs. Enzymes were tested for ADH activity at increasing concentrations of Tyr (0-8 mM) and were normalized to the 0 mM assay. Bars are average activity  $\pm \text{SEM}$  for  $n=3$ . Open symbols correspond to wild-type enzymes, with dashed curves. Mutant enzymes have filled symbols with solid curves.

FIG. 6 shows cofactor specificity of legume noncanonical ADH enzymes. ADH activity was measured for purified recombinant ncADHs from soybean (GmncADH) and *M. truncatula* (MtncADH) using either NADP<sup>+</sup> (gray) or NAD<sup>+</sup>

(black). Bars are average specific activity (nkat/mg)  $\pm \text{SEM}$  ( $n=3$ ). The ratio of ADH activity with NADP<sup>+</sup> to NAD<sup>+</sup> is shown above the bars.

FIGS. 7A-7B show the biochemical characterization of peanut and tomato noncanonical ADH/PDHs. FIG. 7A is a bar graph showing PDH (blue; top bar) and ADH (red; bottom bar) activity of purified recombinant *A. ipaensis* (peanut PDH/ADH) and tomato (SolyncADH) enzymes with NADP<sup>+</sup>. Bars represent average catalytic efficiency ( $k_{cat}/K_m$ ) expressed as  $\text{mM}^{-1} \pm \text{SEM}$  of  $n \geq 3$ . N.D., below detection limit. FIG. 7B is a graph showing the effect of Tyr on ADH (red) and PDH (blue) activity. Data are shown as IC<sub>50</sub> plots with enzymatic activity determined at increasing concentrations of L-Tyr (0-8 mM). Activity was normalized to an assay with no L-Tyr and expressed as percent activity of  $n=3 \pm \text{SEM}$ . Only effects of Tyr on ADH activity from SolyncADH are shown, as it had no activity with prephenate.

FIG. 8 shows the full amino acid sequence alignment of ADH and PDH homologs (SEQ ID NOS: 169-200). Amino acid sequences used in the phylogeny from FIG. 1B were aligned using ClustalW and shaded using BoxShade. Identical residues that are >50% conserved are shaded black, while biochemically similar residues conserved in >50% of the sequences shaded gray. Key catalytic residues are shown in blue (e.g. Ser101, His124, and His188). The cofactor binding domain is highlighted in blue (GxGxxG), with the NAD(P)(H) discriminator region<sup>16</sup> also boxed in blue. From this study, all plant ADH/PDH enzymes are predicted to have NADP(H) specificity, which has been experimentally verified here and previously<sup>5</sup>.  $\beta 1\text{e}-\beta 1\text{f}$  region is highlighted by a gray bar. Asn222 in GmPDH1 is an Asp in all plant ADHs, whereas Met219 in GmPDH1, which is not 100% conserved in ADHs are shaded in red. As in FIG. 1B, blue bars (top nine sequences) represent enzymes with PDH activity and red bars (bottom sequences) represent enzymes with ADH activity. All numbering is based off the GmPDH1 sequence. The sequences are in the order of the phylogeny in FIG. 1B and accession numbers are from the corresponding database where sequences were obtained Phytozome (phytozome.net) and 1KP (onekp.com). Sequence abbreviations, Ad, *Arachis duranensis*; Ai, *Arachis ipaensis*; Am, *Astragalus membranaceus*; At, *Arabidopsis thaliana*; Bb, *Bituminosa bituminaria*; Fv, *Fragaria vesca*; Gg, *Glycyrrhiza glabra*; Gm, *Glycine max*; Gr, *Gossypium raimondii*; Mt, *Medicago truncatula*; Pv, *Phaseolus vulgaris*; Sl, *Solanum lycopersicum*; Tc, *Theobroma cacao*.

FIGS. 9A-9B show the extended phylogenetic analysis of plant TyrA homologs and distribution in Leguminosae. FIG. 9A shows a neighbor-joining phylogenetic analysis created in MEGA6<sup>45</sup> similar to FIG. 1B except with ADH and PDH homologs mainly from legumes. The tree was constructed with 1000 bootstrap values and evolutionary distances were computed using the Poisson correction method involving 90 amino acid sequences. All positions with less than 70% site coverage were eliminated. The noncanonical TyrA clade is shaded gray, stars represent enzymes that were biochemically characterized in this study. FIG. 9B shows the TyrA homolog distribution within the Leguminosae. Presence of TyrA homologs for legumes with sequencing data available were mapped onto a representative Leguminosae taxonomic tree<sup>41,42</sup> with major subclades indicated by black circles. Presence of TyrA homolog is indicated by a filled box (red, canonical or noncanonical ADH, blue PDH) absence is indicated by an empty box. Although limited legume sequences are available, our results suggest that PDHs

duplicated within Leguminosae at least as early as the divergence of Genistoids (*Lupinus* containing) from Dalbergioids (peanut containing).

FIGS. 10A-10C show Asp222 is conserved in plant ADHs and bacterial orthologs. A sequence similarity network<sup>58</sup> was created using GmPDH1 to identify 318 homologs (BLAST® alignment tool  $e\text{-value} \leq 10^{-5}$ ) and visualized in Cytoscape<sup>59</sup>. Each circle (node) represents a single TyrA homolog with each line (edge) connecting the nodes representing two proteins that have sequence similarity greater than a given threshold. FIG. 10A is a pictorial in which 100% networks are shown with increasing sequence similarity scores from left to right of  $\geq 20$ , 25, and 30, respectively. In FIG. 10B the 100% network shows that plant TyrAs (green) are separate, but more closely related to bacterial (red) than archaeal (blue) enzymes. The corresponding residue at position 222 is shown for selected TyrA homologs on top of the node that it represents. *Phenylobacterium zucineum* ( $\alpha$ -proteobacteria ortholog) is from the same genus as *Phenylobacterium immobile* that contains ADH activity<sup>26</sup>. Algal orthologs fall into the plant group including *Cyanidioschyzon merolae* (red algae), *Aureococcus anophagefferens* (brown algae) and *Craspedia variabilis* (green algae), which is from the same genus that contains ADH activity<sup>60</sup>. FIG. 10C shows a trimmed sequence alignment of the TyrA homologs that are marked in panel B showing the corresponding 222 residue (SEQ ID NOS: 201-247).

FIGS. 11A-11C show a structural comparison of plant PDH, cyanobacterial ADH, and bacterial PDH. FIG. 11A shows ribbon diagrams shown as cylinders of GmPDH1 (white, left), SynADH (purple, center), and AaPDH (gold, right) with NAD<sup>+</sup>/NADP<sup>+</sup> (green) shown as a stick model. FIG. 11B shows the NAD<sup>+</sup>/NADP<sup>+</sup> binding sites of GmPDH1, SynADH, and AaPDH show variation in the cis- vs. trans-conformations. The SynADH structure from the PDB depicts the diphosphate moiety in two cis-conformations. Coloring of the ribbons and side-chains is the same as for panel A. FIG. 11C shows the active site residues in GmPDH1 in contact with Tyr (purple), apo SynADH, and AaPDH with 4-hydroxyphenylpyruvate bound (gray).

FIG. 12 shows the conserved acidic residue at 222 among clade I TyrA orthologs from plants, algae, and closely-related bacteria in a structure-guided phylogenetic analysis of plant and microbial TyrAs. Three distinct clades are formed; clade I contains all plant TyrAs and closely-related microbes (blue; top shaded square), clade II contains bacteria, archaea, and fungi TyrAs (green; middle shaded square), and clade III (unshaded at bottom), which was used as an outgroup. Enzymes characterized in this study are marked by black arrows. Structures used to guide the alignment are labeled with their PDB IDs. Previously characterized TyrAs are labeled in red with their preferred PDH or ADH activity. Scale bar represents number of substitutions per branch length. A trimmed amino acid alignment of corresponding sequences shows a conserved acidic residue (Asp or Glu, highlighted in blue) among clade I, which is replaced with a non-acidic Asn or Gln residue (highlighted in green) in most clade II (See SEQ ID NOS: 121-166). Identical amino acids present in >50%, black shading; biochemically similar residues present in >50% of the sequences, gray shading.

FIGS. 13A-13C show substrate and cofactor specificity of microbial TyrA orthologs. ADH and PDH assays were performed with 0.8 mM arogenate and prephenate, respectively, and 0.8 mM cofactor (NADP<sup>+</sup>, black; NAD<sup>+</sup>, gray). FIG. 13A is a bar graph showing purified recombinant

SsTyrA (spirocheates) used to test enzymatic activity, and shown as the average in nKat/mg protein $\pm$ SEM of n=3. FIG. 13B is a bar graph showing  $\alpha$ -proteobacteria TyrA (OiTyrA) cell lysate used as purification of the recombinant enzyme was not successful. Average enzymatic activity is shown as pKat/mg protein $\pm$ SEM of n=3. FIG. 13C is a bar graph showing purified recombinant MhTyrA (archaea) used to test enzymatic activity, and shown as the average in nKat/mg protein $\pm$ SEM of n=3. N.D. no activity detected. Cofactor preference is indicated by the fold-change over the bars.

FIGS. 14A-14B show a kinetic analysis of MhTyrAp wild-type and Q227E mutant enzymes. Kinetic analysis was performed with MhTyrAp wild-type (filled circle) and Q227E mutant (open square) enzymes using various concentrations of prephenate (FIG. 14A) and arogenate (FIG. 14B). Initial velocity values at each substrate concentration were fit to the Michaelis-Menten equation using Origin software. Kinetic analyses were conducted for MhTyrA wild-type using 3.41  $\mu$ g of purified recombinant enzyme, and 4.56  $\mu$ g and 2.28  $\mu$ g of purified recombinant Q227E using prephenate and arogenate, respectively.

FIG. 15 shows structural conservation of residue 222 among clade I TyrA orthologs. Homology models of AtADH2 (blue), SsTyrAa (red), and MhTyrAp (yellow) show that they contain conserved catalytic residues (e.g. His and Ser, numbering based on GmPDH1 structure, which was used as the template for modeling). All three enzymes have an acidic residue at the active site 222: Asp in AtADH2 and SsTyrA and Gln in MhTyrA.

FIG. 16 shows purification of MhTyrAp wild-type (Wt) and Q227E recombinant enzymes. SDS-PAGE of supernatants and recombinant MhTyrApWt and Q227E purified using affinity chromatography facilitated by a 6 $\times$ -His tag on the N-terminus of the protein. *E. coli* supernatants (lanes 1 & 3) expressing MhTyrA Wt and Q227E were applied to a column containing Ni-NTA resin and eluted with 500 mM imidazole containing buffer. Purified recombinant MhTyrA Wt (lane 2) and Q227 (lane 4) eluted at the appropriate size of ~34 kDa.

FIG. 17 shows cofactor specificity of MhTyrAp Q227E mutant. Using the preferred substrate (wild-type (Wt), prephenate, Q227E arogenate) cofactor specificity was tested with NADP<sup>+</sup>(black) and NAD<sup>+</sup>(gray). Mutation of MhTyrAp had not effect on its cofactor preference.

FIG. 18 shows phylogenetic analysis of Spirocheate TyrA orthologs. TyrA orthologs from Spirocheates were identified through BlastP® alignment tool searches using characterized Spirocheates TyrA (SsADH) targeting specific Spirocheates orders (Leptospirales, Brevinematales, and Brachyspirales) that were not included in FIG. 12. TyrA orthologs were identified in the Spirocheates, Leptospirales, and Brachyspirales, but not in Brevinematales. Neighbor-joining phylogenetic analysis performed in MEGA7 from the MUSCLE alignment of Spirocheate TyrA orthologs. Evolutionary distances were calculated using the Poisson correction method with 1,000 bootstrap replicates, which are indicated at the branches, with values less than 50% removed for clarity. Scale bar represents number of amino acid substitutions per site. TyrA orthologs from Spirocheates form a clade with SsADH and plant TryAs (characterized enzymes from this study or in previous studies shown in red). Whereas TyrA from Leptospirales and Brachyspirales group distinctly from clade I, suggesting that only a portion of Spirocheate have plant-like TyrA enzymes that group within clade I. Full genus and species followed by NCBI accession number are indicated for Spirocheate sequences not included in the original phylogenetic analyses (FIG. 12).

FIGS. 19A-19B show conservation of global conformation in divergent microbial TyrA orthologs. One representative sequence from the outgroup (*Bifidobacterium dentium*, BdTyrA) was chosen to determine active site architecture conservation in divergent microbial TyrAs. Models for BdTyrA (red), were created in SWISS-MODEL using GmPDH1 (light red; BdTyrA (GmPDH1)) and a more similar sequence from *Synechocystis* (dark red; BdTyrA (*Synechocystis* ADH)) as templates. FIG. 19A is an overlay showing both BdTyrA models and their template structures. The overall conformation is generally conserved across divergent TyrAs, with some exceptions highlighted with arrows. An extended loop region is present in both models of BdTyrA and *Synechocystis* ADH, and there are additional  $\alpha$ -helices in BdTyrA (*Synechocystis*) and *Synechocystis* ADH. FIG. 19B shows that all enzymes possess the catalytic His and Ser residues, although His112 in *Synechocystis* ADH is in a slightly different position within the active site. The substrate specificity determining residue is present in only GmPDH1 (Asn222), whereas Asp227 in BdTyrA (GmPDH1) is shown but did not align with Asn222 in PROMALS3D alignments and adopts a different conformation and position than Asn222. In BdTyrA (*Synechocystis* ADH) and *Synechocystis* ADH a corresponding residue is lacking entirely in the active site.

#### SUMMARY

In one aspect of the present invention, engineered prephenate dehydrogenases (PDH) and arogenate dehydrogenase/prephenate dehydrogenases (ADH/PDH) polypeptides that have increased ADH activity and tyrosine (Tyr) sensitivity are provided. The engineered prephenate dehydrogenase polypeptides or arogenate dehydrogenase/prephenate dehydrogenase (ADH/PDH) polypeptides may include an aspartic acid (D) amino acid residue or a glutamic acid (E) amino acid residue at a position corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D).

In another aspect, engineered arogenate dehydrogenase (ADH) polypeptides that have increased PDH activity and are less sensitive to tyrosine (Tyr) inhibition are provided. The engineered arogenate dehydrogenase polypeptides may include a non-acidic amino acid residue at a position corresponding to amino acid residue 220 of SEQ ID NO: 10 (MtncADH D220C).

In a further aspect, polynucleotides encoding any one of the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides disclosed herein are provided.

In another aspect, constructs are provided. The constructs may include a promoter operably linked to any one of the polynucleotides described herein.

In a further aspect, vectors including any of the constructs or polynucleotides described herein are provided.

In another aspect, cells including any of the polynucleotides, constructs, or vectors described herein are provided.

In a further aspect, plants including any of the polynucleotides, constructs, vectors, or cells described herein are also provided.

In a still further aspect, methods for increasing production of at least one product of the tyrosine or HPP pathways in a cell are provided. The methods may include introducing any of the polynucleotides, constructs, or vectors described herein into the cell. Optionally, the methods may further include purifying the product of the tyrosine or HPP pathways from the cells.

#### DETAILED DESCRIPTION

Here, the present inventors used phylogeny-guided structure-function analyses of ADHs from legumes and eudicots

that are phylogenetically related to legume PDHs and identified an active site residue (i.e. the amino acid residue at position 220 of SEQ ID NO: 1 (MtPDH C220D) and the corresponding position in other ADH and PDH polypeptides) that determines prephenate versus arogenate specificity in these enzymes and simultaneously alters Tyr feedback inhibition. The structures of mutant PDH enzymes co-crystallized with Tyr reveal the molecular basis of TyrA substrate specificity and feedback-regulation that underlies the evolution of two alternative Tyr pathways in plants. Subsequent mutagenesis of the corresponding residue in divergent plant ADHs introduced PDH activity and relaxed Tyr sensitivity, highlighting the critical role of this residue in TyrA substrate specificity underlying the evolution of alternative Tyr biosynthetic pathways in plants.

In one aspect of the present invention, engineered prephenate dehydrogenase (PDH) polypeptides and arogenate dehydrogenase/prephenate dehydrogenase (ADH/PDH) polypeptides that have increased ADH activity and tyrosine (Tyr) sensitivity are provided. The engineered prephenate dehydrogenase polypeptides or arogenate dehydrogenase/prephenate dehydrogenase (ADH/PDH) polypeptides may include an aspartic acid (D) amino acid residue or a glutamic acid (E) amino acid residue at a position corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D).

The engineered PDH polypeptides or ADH/PDH polypeptides may include a polypeptide or a functional fragment thereof having at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 98%, 99%, or 100% sequence identity to any one of the polypeptides of SEQ ID NOS: 1-9, 121-123, 144-148, 152-158, 213-217, or 243-247 and including an aspartic acid (D) amino acid residue or a glutamic acid (E) amino acid residue at a position corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D).

As used herein, the phrase "at a position corresponding to" refers to an amino acid position that aligns with an amino acid position of another identified sequence in a protein sequence alignment or a protein structure alignment. For example, the phrase "at a position corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D)" refers to an amino acid position in a polypeptide sequence that aligns with the 220<sup>th</sup> amino acid residue in SEQ ID NO: 1 (MtPDH C220) when the two polypeptide sequences are aligned using common sequence alignment programs. Regarding SEQ ID NOS: 1-55 and 121-158, the amino acid positions in these polypeptide sequences corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D) are shown as the rightmost asterisk in the partial sequence alignment shown in FIG. 3A and as the asterisk in FIG. 12. SEQ ID NOS: 1-55 represent engineered versions of the polypeptides represented in FIG. 3A. SEQ ID NOS: 121-158 represent non-engineered versions of the polypeptides represented in FIG. 12. Thus, SEQ ID NOS: 1-9 are the top nine PDH and PDH/ADH polypeptides shown in the partial sequence alignment in FIG. 3A where the asparagine (N) or cysteine (C) amino acid residue at the position corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D) (the asterisk labeled N222) is substituted with an aspartic acid (D) amino acid residue. SEQ ID NOS: 121-123, 144-148, and 152-158 are the PDH and PDH/ADH polypeptides shown in the partial sequence alignment in FIG. 12. SEQ ID NOS: 10-55 represent the bottom 23 ADH polypeptides shown in the partial sequence alignment in FIG. 3A where the aspartic acid (D) amino acid residue at the position corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D) (the asterisk labeled N222) (also identified as a position corresponding to amino acid

residue 220 of SEQ ID NO: 10 (MtnCADH D220C)) is substituted with either an asparagine (N) amino acid residue or a cysteine (C) amino acid residue. SEQ ID NOs: 124-143 and 149-151 are the ADH polypeptides shown in the partial sequence alignment in FIG. 12.

To determine whether a particular polypeptide sequence has an amino acid residue position "corresponding to" an identified sequence disclosed herein, a person of ordinary skill may align the particular sequence with the sequences described in FIG. 12 (SEQ ID NOs: 121-166) using the methods described in FIG. 12. See, e.g., Pei et al., PRO-MALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* 23(7): 802-8 (2007). If the particular sequence falls within clades I or II (SEQ ID NOs: 121-158), then the particular sequence does have an amino acid residue corresponding to the identified sequence disclosed herein, which can be determined by examining the sequence alignment at the appropriate position. If, however, the particular sequence falls within clade III (SEQ ID NOs: 159-166), then the particular sequence does not have an amino acid residue corresponding to the identified sequence disclosed herein.

In the Examples, the present inventors demonstrated that the polypeptides of SEQ ID NOs: 1 and 2 demonstrated a switch in substrate specificity from primarily PDH activity to primarily ADH activity and also introduced Tyr sensitivity into the enzymes. Likewise, the present inventors expect that the polypeptides of SEQ ID NOs: 3-9 would also exhibit increased ADH activity and Tyr sensitivity and that the polypeptides of SEQ ID NOs: 1-9, 121-123, 144-148, 152-158, 213-217, and 243-247, when engineered to include an aspartic acid (D) amino acid residue or a glutamic acid (E) amino acid residue at a position corresponding to amino acid residue 220 of SEQ ID NO: 1 (MtPDH C220D), may also exhibit increased ADH activity and Tyr sensitivity. Thus, in some embodiments, the engineered prephenate dehydrogenases (PDH) and arogenate dehydrogenase/prephenate dehydrogenases (ADH/PDH) polypeptides disclosed herein may have greater arogenate dehydrogenase activity than prephenate dehydrogenase activity. In some embodiments, the arogenate dehydrogenase activity of the engineered prephenate dehydrogenases (PDH) and arogenate dehydrogenase/prephenate dehydrogenase (ADH/PDH) polypeptides may be 1.5, 2, 3, 5, 10, 20, or more fold greater than the prephenate dehydrogenase activity.

As used herein, a polypeptide may "have greater arogenate dehydrogenase activity than prephenate dehydrogenase activity" or "have greater prephenate dehydrogenase activity than arogenate dehydrogenase activity" when the steady-state kinetic parameters ( $k_{cat}/K_m$  ( $\text{mM}^{-1} \text{s}^{-1}$ )) for arogenate dehydrogenase activity are greater than the steady-state parameters ( $k_{cat}/K_m$  ( $\text{mM}^{-1} \text{s}^{-1}$ )) for prephenate dehydrogenase activity or when the steady-state kinetic parameters ( $k_{cat}/K_m$  ( $\text{mM}^{-1} \text{s}^{-1}$ )) for prephenate dehydrogenase activity are greater than the steady-state parameters ( $k_{cat}/K_m$  ( $\text{mM}^{-1} \text{s}^{-1}$ )) for arogenate dehydrogenase activity. Steady-state kinetic parameters may be measured using techniques similar to those described by the inventors in the Examples. Briefly, kinetic parameters of purified polypeptides can be determined from assays conducted at varying arogenate and prephenate concentrations. Standard assay conditions include 25 mM HEPES pH 7.6, 50 mM KCl and 10% (v/v) ethylene glycol, and 0.5 mM NADP<sup>+</sup> with varied substrate, concentrations. Reactions can be initiated by the addition of the polypeptide and incubated at 37° C. monitored every 10-15 seconds at A<sub>340nm</sub> using a microplate reader. Kinetic

parameters may be determined by fitting initial velocity data to the Michaelis-Menten equation using the Origin software.

In some embodiments, the engineered prephenate dehydrogenases (PDH) and arogenate dehydrogenase/prephenate dehydrogenase (ADH/PDH) polypeptides may include SEQ ID NO: 1 (MtPDH C220D), SEQ ID NO: 2 (GmPDH1 N222D), a polypeptide having at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 98%, 99% sequence identity to SEQ ID NO: 1 and including an aspartic acid (D) amino acid residue or a glutamic acid (E) residue at position 220 of SEQ ID NO: 1, or a polypeptide having at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 98%, 99% sequence identity to SEQ ID NO: 2 and including the aspartic acid (D) amino acid residue or a glutamic acid (E) residue at position 222 of SEQ ID NO: 2.

In some embodiments, the engineered prephenate dehydrogenases (PDH) and arogenate dehydrogenase/prephenate dehydrogenase (ADH/PDH) polypeptides may include SEQ ID NO: 1 (MtPDH C220D) or SEQ ID NO: 2 (GmPDH1 N222D).

In another aspect of the present invention, engineered arogenate dehydrogenase (ADH) polypeptides that have increased PDH activity and are less sensitive to tyrosine (Tyr) inhibition are provided. The engineered arogenate dehydrogenase polypeptides may include a non-acidic amino acid residue at a position corresponding to amino acid residue 220 of SEQ ID NO: 10 (MtnCADH D220C).

As used herein, a "non-acidic" amino acid may include any amino acid except aspartic acid (D) or glutamic acid (E) and may include, without limitation, Alanine (A), Arginine (R), Asparagine (N), Cysteine (C), Glutamine (Q), Glycine (G), Histidine (H), Isoleucine (I), Leucine (L), Lysine (K), Methionine (M), Phenylalanine (F), Proline (P), Serine (S), Threonine (T), Tryptophan (W), Tyrosine (Y), or Valine (V). In some embodiments, the non-acidic amino acid residue may be an asparagine (N) amino acid residue or a cysteine (C) amino acid residue.

The engineered ADH polypeptides may include a polypeptide or a functional fragment thereof having at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 98%, 99% sequence identity to any one of the polypeptides of SEQ ID NOs: 10-55, 124-143, 149-151 201-212, or 218-242 and including a non-acidic amino acid residue at a position corresponding to amino acid residue 220 of SEQ ID NO: 10 (MtnCADH D220C).

The engineered ADH polypeptides may have greater prephenate dehydrogenase activity than arogenate dehydrogenase activity. In some embodiments, the prephenate dehydrogenase activity of the engineered ADH polypeptides may be 1.5, 2, 3, 5, 10, 20, or more fold greater than the arogenate dehydrogenase activity.

In some embodiments, the engineered ADH polypeptide may include SEQ ID NO: 10 (MtnCADH D220C), SEQ ID NO: 11 (MtnCADH D220N), SEQ ID NO: 12 (AtADH2 D241N), SEQ ID NO: 13 (AtADH2 D241C), a polypeptide having at least 80% sequence identity to SEQ ID NO: 10 and including a cysteine (C) amino acid residue at position 220 of SEQ ID NO: 10, a polypeptide having at least 80% sequence identity to SEQ ID NO: 11 and including an asparagine (N) amino acid residue at position 220 of SEQ ID NO: 11, a polypeptide having at least 80% sequence identity to SEQ ID NO: 12 and including an asparagine (N) amino acid residue at position 241 of SEQ ID NO: 12, and a polypeptide having at least 80% sequence identity to SEQ ID NO: 13 and including a cysteine (C) amino acid residue at position 241 of SEQ ID NO: 13.

In some embodiments, the engineered ADH polypeptides may include any one of the polypeptides of SEQ ID NOs: 10-13.

The engineered ADH polypeptides having PDH activity may also not be sensitive to tyrosine inhibition. The polypeptide is considered to not be sensitive, i.e. to lack sensitivity to tyrosine feedback inhibition if at least 80% of the activity of the polypeptide in the absence of tyrosine is maintained in the presence of 1 mM tyrosine.

Regarding the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides disclosed herein, the phrases “% sequence identity,” “percent identity,” or “% identity” refer to the percentage of residue matches between at least two amino acid sequences aligned using a standardized algorithm. Methods of amino acid sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail below, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. Percent identity for amino acid sequences may be determined as understood in the art. (See, e.g., U.S. Pat. No. 7,396,664, which is incorporated herein by reference in its entirety). A suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST® alignment tool), which is available from several sources, including the NCBI, Bethesda, Md., at its website. The BLAST® alignment tool software suite includes various sequence analysis programs including “blastp,” that is used to align a known amino acid sequence with other amino acid sequences from a variety of databases.

Polypeptide sequence identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

The engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides disclosed herein may include “variant” polypeptides, “mutants,” and “derivatives thereof.” As used herein, a “variant,” “mutant,” or “derivative” refers to a polypeptide molecule having an amino acid sequence that differs from a reference protein or polypeptide molecule. A variant or mutant may have one or more insertions, deletions, or substitutions of an amino acid residue relative to a reference molecule. For example, an engineered PDH, PDH/ADH, ADH dehydrogenase polypeptide mutant or variant may have one or more insertion, deletion, or substitution of at least one amino acid residue relative to the reference engineered PDH, PDH/ADH, ADH dehydrogenase polypeptides disclosed herein. The polypeptide sequences of the engineered PDH, PDH/ADH, ADH dehydrogenase polypeptides from various species are presented in SEQ ID NOs: 1-55 and 121-158. These sequences may be used as reference sequences.

The engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides provided herein may be full-length polypeptides or may be fragments of the full-length polypeptide. As used herein, a “fragment” is a portion of an amino acid

sequence which is identical in sequence to, but shorter in length than a reference sequence. A fragment may comprise up to the entire length of the reference sequence, minus at least one amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous amino acid residues of a reference polypeptide, respectively. In some embodiments, a fragment may comprise at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 250, or 500 contiguous amino acid residues of a reference polypeptide. Fragments may be preferentially selected from certain regions of a molecule. The term “at least a fragment” encompasses the full-length polypeptide. A fragment of an ADH polypeptide may comprise or consist essentially of a contiguous portion of an amino acid sequence of the full-length ADH polypeptide (See, e.g., SEQ ID NOs: 1-55, 121-158, 201-247). A fragment may include an N-terminal truncation, a C-terminal truncation, or both truncations relative to the full-length ADH polypeptide.

A “deletion” in an engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptide refers to a change in the amino acid sequence resulting in the absence of one or more amino acid residues. A deletion may remove at least 1, 2, 3, 4, 5, 10, 20, 50, 100, 200, or more amino acids residues. A deletion may include an internal deletion and/or a terminal deletion (e.g., an N-terminal truncation, a C-terminal truncation or both of a reference polypeptide).

“Insertions” and “additions” in an engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptide refers to changes in an amino acid sequence resulting in the addition of one or more amino acid residues. An insertion or addition may refer to 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more amino acid residues. A variant of an engineered PDH, PDH/ADH, ADH dehydrogenase polypeptide may have N-terminal insertions, C-terminal insertions, internal insertions, or any combination of N-terminal insertions, C-terminal insertions, and internal insertions.

The amino acid sequences of the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptide variants, mutants, derivatives, or fragments as contemplated herein may include conservative amino acid substitutions relative to a reference amino acid sequence. For example, a variant, mutant, derivative, or fragment polypeptide may include conservative amino acid substitutions relative to a reference molecule. “Conservative amino acid substitutions” are those substitutions that are a substitution of an amino acid for a different amino acid where the substitution is predicted to interfere least with the properties of the reference polypeptide. In other words, conservative amino acid substitutions substantially conserve the structure and the function of the reference polypeptide. Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

The disclosed variant and fragment engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides described herein may have one or more functional or biological activities exhibited by a reference polypeptide (i.e., SEQ ID NOs: 1-55 or engineered versions of SEQ ID NOs: 121-158). Suitably, the disclosed variant or fragment engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides retain at least 20%, 40%, 60%, 80%, or 100% of the argonate dehydrogenase activity or the prephenate dehydrogenase activity of the reference polypeptide (i.e., SEQ ID NOs: 1-55 or engineered versions of SEQ ID NOs: 121-158 or 201-247).

As used herein, a “functional fragment” of an engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptide is a fragment of, for example, one of the polypeptides of SEQ ID NOS: 1-15 that retains at least 20%, 40%, 60%, 80%, or 100% of the arogenate dehydrogenase activity or the prephenate dehydrogenase activity of the full-length polypeptide. Exemplary functional fragments of the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides disclosed herein may include, for example, the highly-conserved amino acid residues responsible for NADP<sup>+</sup> binding, including the GxGxxG motif, and residues proposed to function in catalysis (e.g. Ser101 and His124). See FIG. 8.

FIG. 8 shows a sequence alignment including the PDH, PDH/ADH, and ADH dehydrogenase polypeptides, which were engineered and disclosed as SEQ ID NOS: 1-55. Based on this alignment it becomes immediately apparent to a person of ordinary skill in the art that various amino acid residues may be altered (i.e. substituted, deleted, etc.) without substantially affecting the arogenate dehydrogenase activity or the prephenate dehydrogenase activity of the polypeptide. For example, a person of ordinary skill in the art would appreciate that substitutions in a reference PDH, PDH/ADH, or ADH dehydrogenase polypeptide could be based on alternative amino acid residues that occur at the corresponding position in other PDH, PDH/ADH, or ADH dehydrogenase polypeptide from other species. For example, the MtPDH polypeptide in FIG. 8 has a threonine amino acid residue at position 57 while some of the other polypeptides in FIG. 8 have a serine, alanine, or other amino acid at this position in the alignment. Thus, one exemplary modification that is apparent from the sequence alignment in FIG. 8 is a T57S or T57A substitution in the disclosed engineered MtPDH polypeptide (SEQ ID NO: 1). Similar modifications could be made to each of SEQ ID NOS: 1-55 at each position of the sequence alignment shown in FIG. 8. Additionally, a person of ordinary skill in the art could easily align other PDH, PDH/ADH, ADH dehydrogenase polypeptides with the polypeptide sequences shown in FIG. 8 to determine what additional variants could be made to the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides.

The engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides contemplated herein may be further modified in vitro or in vivo to include non-amino acid moieties. These modifications may include but are not limited to acylation (e.g., O-acylation (esters), N-acylation (amides), S-acylation (thioesters)), acetylation (e.g., the addition of an acetyl group, either at the N-terminus of the protein or at lysine residues), formylation, lipoylation (e.g., attachment of a lipoate, a C8 functional group), myristoylation (e.g., attachment of myristate, a C14 saturated acid), palmitoylation (e.g., attachment of palmitate, a C16 saturated acid), alkylation (e.g., the addition of an alkyl group, such as an methyl at a lysine or arginine residue), isoprenylation or prenylation (e.g., the addition of an isoprenoid group such as farnesol or geranylgeraniol), amidation at C-terminus, glycosylation (e.g., the addition of a glycosyl group to either asparagine, hydroxylysine, serine, or threonine, resulting in a glycoprotein). Distinct from glycation, which is regarded as a nonenzymatic attachment of sugars, polysialylation (e.g., the addition of polysialic acid), glypiation (e.g., glycosylphosphatidylinositol (GPI) anchor formation, hydroxylation, iodination (e.g., of thyroid hormones), and phosphorylation (e.g., the addition of a phosphate group, usually to serine, tyrosine, threonine or histidine) are enzymatic or covalent attachments.

Polynucleotides encoding any one of the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides disclosed herein are provided. As used herein, the terms “polynucleotide,” “polynucleotide sequence,” “nucleic acid” and “nucleic acid sequence” refer to a nucleotide, oligonucleotide, polynucleotide (which terms may be used interchangeably), or any fragment thereof. These phrases also refer to DNA or RNA of natural or synthetic origin (which may be single-stranded or double-stranded and may represent the sense or the antisense strand). The polynucleotides may be cDNA or genomic DNA.

Polynucleotides homologous to the polynucleotides described herein are also provided. Those of skill in the art understand the degeneracy of the genetic code and that a variety of polynucleotides can encode the same polypeptide. In some embodiments, the polynucleotides (i.e., polynucleotides encoding the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides) may be codon-optimized for expression in a particular cell including, without limitation, a plant cell, bacterial cell, or fungal cell. While particular polynucleotide sequences which are found in plants are disclosed herein any polynucleotide sequences may be used which encode a desired form of the polypeptides described herein. The particular polynucleotide sequences of the non-engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides are provided as SEQ ID NOS: 56-96. Thus non-naturally occurring sequences may be used. These may be desirable, for example, to enhance expression in heterologous expression systems of polypeptides or proteins. Computer programs for generating degenerate coding sequences are available and can be used for this purpose. Pencil, paper, the genetic code, and a human hand can also be used to generate degenerate coding sequences.

In another aspect of the present invention, constructs are provided. As used herein, the term “construct” refers to recombinant polynucleotides including, without limitation, DNA and RNA, which may be single-stranded or double-stranded and may represent the sense or the antisense strand. Recombinant polynucleotides are polynucleotides formed by laboratory methods that include polynucleotide sequences derived from at least two different natural sources or they may be synthetic. Constructs thus may include new modifications to endogenous genes introduced by, for example, genome editing technologies. Constructs may also include recombinant polynucleotides created using, for example, recombinant DNA methodologies.

The constructs provided herein may be prepared by methods available to those of skill in the art. Notably each of the constructs claimed are recombinant molecules and as such do not occur in nature. Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, and recombinant DNA techniques that are well known and commonly employed in the art. Standard techniques available to those skilled in the art may be used for cloning, DNA and RNA isolation, amplification and purification. Such techniques are thoroughly explained in the literature.

The constructs provided herein may include a promoter operably linked to any one of the polynucleotides described herein. The promoter may be a heterologous promoter or an endogenous promoter associated with the PDH, PDH/ADH, or ADH dehydrogenase polypeptide.

As used herein, the terms “heterologous promoter,” “promoter,” “promoter region,” or “promoter sequence” refer generally to transcriptional regulatory regions of a gene, which may be found at the 5' or 3' side of the ADH polynucleotides described herein, or within the coding

region of the ADH polynucleotides, or within introns in the ADH polynucleotides. Typically, a promoter is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. The typical 5' promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site (conveniently defined by mapping with nuclease Si), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

In some embodiments, the disclosed PDH, PDH/ADH, or ADH dehydrogenase polynucleotides are operably connected to the promoter. As used herein, a polynucleotide is "operably connected" or "operably linked" when it is placed into a functional relationship with a second polynucleotide sequence. For instance, a promoter is operably linked to a PDH, PDH/ADH, or ADH dehydrogenase polynucleotide if the promoter is connected to the PDH, PDH/ADH, or ADH dehydrogenase polynucleotide such that it may effect transcription of the PDH, PDH/ADH, or ADH dehydrogenase polynucleotides. In various embodiments, the PDH, PDH/ADH, or ADH dehydrogenase polynucleotides may be operably linked to at least 1, at least 2, at least 3, at least 4, at least 5, or at least 10 promoters.

Heterologous promoters useful in the practice of the present invention include, but are not limited to, constitutive, inducible, temporally-regulated, developmentally regulated, chemically regulated, tissue-preferred and tissue-specific promoters. The heterologous promoter may be a plant, animal, bacterial, fungal, or synthetic promoter. Suitable promoters for expression in plants include, without limitation, the 35S promoter of the cauliflower mosaic virus, ubiquitin, tCUP cryptic constitutive promoter, the Rsyn7 promoter, pathogen-inducible promoters, the maize In2-2 promoter, the tobacco PR-1a promoter, glucocorticoid-inducible promoters, estrogen-inducible promoters and tetracycline-inducible and tetracycline-repressible promoters. Other promoters include the T3, T7 and SP6 promoter sequences, which are often used for in vitro transcription of RNA. In mammalian cells, typical promoters include, without limitation, promoters for Rous sarcoma virus (RSV), human immunodeficiency virus (HIV-1), cytomegalovirus (CMV), SV40 virus, and the like as well as the translational elongation factor EF-1 $\alpha$  promoter or ubiquitin promoter. Those of skill in the art are familiar with a wide variety of additional promoters for use in various cell types. In some embodiments, the heterologous promoter includes a plant promoter, either endogenous to the plant host or heterologous.

Vectors including any of the constructs or polynucleotides described herein are provided. The term "vector" is intended to refer to a polynucleotide capable of transporting another polynucleotide to which it has been linked. In some embodiments, the vector may be a "plasmid," which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome, such as some viral vectors or transposons. Plant mini-chromosomes are also included as

vectors. Vectors may carry genetic elements, such as those that confer resistance to certain drugs or chemicals.

Cells including any of the polynucleotides, constructs, or vectors described herein are provided. Suitable "cells" that may be used in accordance with the present invention include eukaryotic or prokaryotic cells. Suitable eukaryotic cells include, without limitation, plant cells, fungal cells, and animal cells. Suitable prokaryotic cells include, without limitation, gram-negative and gram-positive bacterial species. In some embodiments, the cell is a plant cell such as, without limitation, a beet plant cell, a soybean plant cell, a mung bean plant cell, an opium poppy plant cell, an alfalfa plant cell, a rice plant cell, a wheat plant cell, a corn plant cell, a sorghum plant cell, a barley plant cell, a millet plant cell, an oat plant cell, a rye plant cell, a rapeseed plant cell, and a miscanthus plant cell. In some embodiments, the cell is a bacterial or fungal cell. For example, the polynucleotides, constructs, or vectors described herein may be introduced into yeast cells to improve the production of opioids such as morphine. See, e.g., Galanie et al., DOI: 10.1126/science.aac9373, Published Online Aug. 13, 2015.

Plants including any of the polynucleotides, constructs, vectors, or cells described herein are also provided. Suitable plants may include, without limitation, a beet plant, a soybean plant, a mung bean plant, an opium poppy plant, an alfalfa plant, a rice plant, a wheat plant, a corn plant, a sorghum plant, a barley plant, a millet plant, an oat plant, a rye plant, and a rapeseed plant as well as perennial grasses such as a miscanthus plant. For example, polynucleotides encoding any one of the engineered PDH, PDH/ADH, or ADH dehydrogenase polypeptides of SEQ ID NOs: 1-55 may be used to generate transgenic plants.

Portions or parts of these plants are also useful and provided. Portions and parts of plants includes, without limitation, plant cells, plant tissue, plant progeny, plant asexual propagates, plant seeds. The plant may be grown from a seed comprising transgenic cells or may be grown by any other means available to those of skill in the art. Chimeric plants comprising transgenic cells are also provided and encompassed.

As used herein, a "plant" includes any portion of the plant including, without limitation, a whole plant, a portion of a plant such as a part of a root, leaf, stem, seed, pod, flower, cell, tissue plant germplasm, asexual propagate, or any progeny thereof. Germplasm refers to genetic material from an individual or group of individuals or a clone derived from a line, cultivar, variety or culture. Plant refers to whole plants or portions thereof including, without limitation, plant cells, plant protoplasts, plant tissue culture cells or calli. For example, a soybean plant refers to whole soybean plant or portions thereof including, without limitation, soybean plant cells, soybean plant protoplasts, soybean plant tissue culture cells or calli. A plant cell refers to cells harvested or derived from any portion of the plant or plant tissue culture cells or calli.

Methods for increasing production of at least one product of the tyrosine or HPP pathways in a cell are provided. The methods may include introducing any of the polynucleotides, constructs, or vectors described herein into the cell. Suitable products of the tyrosine or HPP pathways include, without limitation, vitamin E, plastoquinone, a cyanogenic glycoside, a benzyloisoquinoline alkaloid, rosmarinic acid, betalains, suberin, mescaline, morphine, salidroside, a phenylpropanoid compound, dhurrin, a tocochromanol, ubiquinone, lignin, a catecholamine such as epinephrine (adrenaline) or dopamine (i.e., L-dihydroxyphenylalanine (L-DOPA)), melanin, an isoquinoline alkaloid, hydroxycin-

amic acid amide (HCAA), an amaryllidaceae alkaloid, hordenine, hydroxycinnamate, hydroxylstyrene, or tyrosine. Phenylpropanoid compounds (i.e., lignin, tannins, flavonoids, stilbene) may be produced from tyrosine, for example, by combining the polypeptides disclosed herein with a tyrosine-ammonia lyase (TAL) or by using cells that naturally have a TAL such as grass cells.

As used herein, “introducing” describes a process by which exogenous polynucleotides (e.g., DNA or RNA) are introduced into a recipient cell. Methods of introducing polynucleotides into a cell are known in the art and may include, without limitation, microinjection, transformation, and transfection methods. Transformation or transfection may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a host cell. The method for transformation or transfection is selected based on the type of host cell being transformed and may include, but is not limited to, the floral dip method, *Agrobacterium*-mediated transformation, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. Microinjection of polynucleotides may also be used to introduce polynucleotides into cells.

In some embodiments, the present methods may further include purifying the product of the tyrosine or HPP pathways from the cells. As used herein, the term “purifying” is used to refer to the process of ensuring that the product of the tyrosine or HPP pathways is substantially or essentially free from cellular components and other impurities. Purification of products of the tyrosine or HPP pathways is typically performed using analytical chemistry techniques such as high performance liquid chromatography and other chromatographic techniques. Methods of purifying such products are well known to those skilled in the art. A “purified” product of the tyrosine or HPP pathways means that the product is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The present disclosure is not limited to the specific details of construction, arrangement of components, or method steps set forth herein. The compositions and methods disclosed herein are capable of being made, practiced, used, carried out and/or formed in various ways that will be apparent to one of skill in the art in light of the disclosure that follows. The phraseology and terminology used herein is for the purpose of description only and should not be regarded as limiting to the scope of the claims. Ordinal indicators, such as first, second, and third, as used in the description and the claims to refer to various structures or method steps, are not meant to be construed to indicate any specific structures or steps, or any particular order or configuration to such structures or steps. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., “such as”) provided herein, is intended merely to facilitate the disclosure and does not imply any limitation on the scope of the disclosure unless otherwise claimed. No language in the specification, and no structures shown in the drawings, should be construed as indicating that any non-claimed element is essential to the practice of the disclosed subject matter. The use herein of the terms “including,” “comprising,” or “having,” and variations thereof, is meant to encompass the elements listed thereafter and equivalents thereof, as well as additional elements. Embodiments recited as “including,” “comprising,” or “hav-

ing” certain elements are also contemplated as “consisting essentially of” and “consisting of” those certain elements.

Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. For example, if a concentration range is stated as 1% to 50%, it is intended that values such as 2% to 40%, 10% to 30%, or 1% to 3%, etc., are expressly enumerated in this specification. These are only examples of what is specifically intended, and all possible combinations of numerical values between and including the lowest value and the highest value enumerated are to be considered to be expressly stated in this disclosure. Use of the word “about” to describe a particular recited amount or range of amounts is meant to indicate that values very near to the recited amount are included in that amount, such as values that could or naturally would be accounted for due to manufacturing tolerances, instrument and human error in forming measurements, and the like. All percentages referring to amounts are by weight unless indicated otherwise.

No admission is made that any reference, including any non-patent or patent document cited in this specification, constitutes prior art. In particular, it will be understood that, unless otherwise stated, reference to any document herein does not constitute an admission that any of these documents forms part of the common general knowledge in the art in the United States or in any other country. Any discussion of the references states what their authors assert, and the applicant reserves the right to challenge the accuracy and pertinence of any of the documents cited herein. All references cited herein are fully incorporated by reference in their entirety, unless explicitly indicated otherwise. The present disclosure shall control in the event there are any disparities between any definitions and/or description found in the cited references.

Unless otherwise specified or indicated by context, the terms “a,” “an,” and “the” mean “one or more.” For example, “a protein” or “an RNA” should be interpreted to mean “one or more proteins” or “one or more RNAs,” respectively.

The following examples are meant only to be illustrative and are not meant as limitations on the scope of the invention or of the appended claims.

## EXAMPLES

### Example 1—Molecular Basis of the Evolution of Alternative Tyrosine Biosynthetic Pathways in Plants

This Example is based on data reported in Schenck et al., “Molecular basis of the evolution of alternative tyrosine biosynthetic routes in plants,” *Nat. Chem. Biol.*, 13(9):1029-1035 (2017), the contents of which (including all supplemental data, figures, and associated materials) is incorporated herein by reference.

L-Tyrosine (Tyr) is essential for protein synthesis and a precursor of numerous specialized metabolites crucial for plant and human health. Tyr can be synthesized via two alternative routes by a key regulatory TyrA family enzyme, prephenate or arogenate dehydrogenase (PDH/TyrA<sub>p</sub> or ADH/TyrA<sub>a</sub>), representing a unique divergence of primary metabolic pathways. However, the molecular foundation underlying the evolution of the alternative Tyr pathways is currently unknown. Here we characterized recently-di-

verged plant PDH and ADHs, obtained the x-ray crystal structure of soybean PDH, and identified a single amino acid residue that defines TyrA substrate specificity and regulation. Structures of mutated PDHs co-crystallized with Tyr indicate that substitutions of Asn222 confers ADH activity and Tyr-sensitivity. Subsequent mutagenesis of the corresponding residue in divergent plant ADHs introduced PDH activity and relaxed Tyr sensitivity, highlighting the critical role of this residue in TyrA substrate specificity underlying the evolution of alternative Tyr biosynthetic pathways in plants.

Unlike recently-evolved and lineage-specific diverse specialized (secondary) metabolic pathways<sup>1</sup>, primary metabolism such as amino acid biosynthesis are ubiquitous and usually conserved among organisms. However, there are some exceptions to this notion<sup>2,3</sup>, and L-tyrosine (Tyr) biosynthetic pathway is one example in which variations have long been described in microbes and plants<sup>4,5</sup>. Elucidation of evolutionary diversification of primary metabolism not only addresses the extent of metabolic plasticity but also provides useful engineering tools to modify core metabolic pathways.

Tyr is an essential aromatic amino acid required for protein synthesis in all organisms but, synthesized de novo only in plants and microorganisms<sup>6,7</sup>. Neurotransmitters such as catecholamines in metazoans are derived from Tyr, which must be obtained from their diet, as they cannot synthesize Tyr de novo<sup>8</sup>. In plants, Tyr serves as the precursor to numerous specialized metabolites crucial for both plant and human health, such as antioxidants vitamin E, the photosynthetic electron carrier plastoquinone, betalain pigments, and defense compounds, including dhurrin, rosmarinic acid, and isoquinoline alkaloids (e.g. morphine)<sup>9-14</sup>. The major plant cell wall component lignin can also be synthesized from Tyr in grasses<sup>15</sup>.

Tyr is synthesized from prephenate, a shikimate pathway product, by two reactions, an oxidative decarboxylation and a transamination (FIG. 1A). The TyrA enzymes catalyze the oxidative decarboxylation step and are the key regulatory enzymes of Tyr biosynthesis, as they are usually inhibited by Tyr and compete for substrates that are also used in L-phenylalanine biosynthesis (FIG. 1A)<sup>16-18</sup>. In many microbes an NAD(H)-dependent prephenate dehydrogenase/TyrA (PDH/TyrA<sub>p</sub>; EC 1.3.1.13) converts prephenate into 4-hydroxyphenylpyruvate (HPP) followed by transamination to Tyr by Tyr aminotransferase (TAT, FIG. 1A)<sup>18</sup>. In plants, these two reactions occur in the reverse order, with prephenate first being transaminated to arogenate by prephenate aminotransferase (PPA-AT), followed by oxidative decarboxylation to Tyr by an NADP(H)-dependent arogenate dehydrogenase/TyrA (ADH/TyrA<sub>a</sub>; EC 1.3.1.78, FIG. 1A)<sup>19-24</sup>. Some exceptions to these “textbook” models are found in nature including microbes that use ADH to synthesize Tyr<sup>25,26</sup> and plants such as legumes having PDH activity<sup>5,27,28</sup>. Also, some microbial TyrAs prefer NADP(H) cofactor<sup>18,29</sup>. Thus, variations exist in the TyrA enzymes in diverse organisms, yet the molecular basis underlying TyrA substrate specificity and the alternative Tyr pathways is currently unknown.

Comparison of microbial TyrA sequences identified an aspartate residue downstream of the NAD(P)(H) binding motif that was later shown to confer cofactor specificity of TyrA<sup>16,30</sup>. Site-directed mutagenesis of *Escherichia coli* PDH and structural analysis of *Aquifex aeolicus* PDH identified an active site histidine, which interacts with substrate C4-hydroxyl and is critical for catalysis in each PDH. The same studies also showed that an active site arginine is

necessary for substrate binding, but not for substrate specificity<sup>31-34</sup>. Besides their varied substrate and cofactor specificities, TyrA enzymes also exhibit different regulatory properties. Mutation of another active site histidine, which is present in the *E. coli* and *A. aeolicus* PDHs but absent in Tyr-insensitive *Synechocystis* ADH, relieved Tyr inhibition but simultaneously reduced PDH activity<sup>34</sup>. Random mutagenesis of the *E. coli* enzyme identified additional residues that relaxed Tyr inhibition; however, PDH activity was also reduced in these mutants<sup>35</sup>. Sequence and structural comparisons of divergent TyrA homologs have been unable to identify specific determinants of Tyr-sensitivity and substrate specificity<sup>1,6,29,30,33,34</sup>.

Recent work described legume PDHs that were insensitive to Tyr regulation<sup>5</sup>. Here, we used phylogeny-guided structure-function analyses of ADHs from legumes and eudicots that are phylogenetically related to legume PDHs and identified an active site residue that determines prephenate versus arogenate specificity in these enzymes and simultaneously alters Tyr inhibition. The structures of mutant PDH enzymes co-crystallized with Tyr reveal the molecular basis of TyrA substrate specificity and feedback-regulation that underlies the evolution of two alternative Tyr pathways in plants.

## Results

### Identification and Biochemical Analysis of Noncanonical ADH in Legumes

Our previous phylogenetic analysis of plant TyrA enzymes (hereafter referred to as either ADH or PDH) identified a “noncanonical” clade (gray box in FIG. 1B) containing legume PDHs that was distinct from the “canonical” ADHs present in all plant lineages<sup>5</sup>. The “noncanonical” clade also contained additional homologs from some eudicots (FIG. 1B). For comparison of the biochemical properties of PDHs and their noncanonical TyrA homologs, representative members of each group were expressed as recombinant proteins and purified for steady-state kinetic analysis and compared with previously characterized canonical ADHs (FIG. 1C; Table 1). PDHs from *Glycine max* (soybean; GmPDH1; 18g02650) and *Medicago truncatula* (MtPDH; 3g071980) preferred prephenate versus arogenate as substrates with 139-fold and 21-fold higher  $k_{cat}/K_m$  values, respectively. The noncanonical TyrA homolog from soybean (Gm14g05990) only displayed activity with arogenate, whereas that from *M. truncatula* (Mt5g083530) accepted both substrates but was 6,200-fold more efficient with arogenate, similar to previously characterized ADH from *Arabidopsis thaliana* (AtADH2; At1g15710)<sup>19</sup>. Thus, Gm14g05990 and Mt5g083530 are noncanonical ADHs (GmncADH and MtncADH, respectively). Each of the legume noncanonical ADH used NADP<sup>+</sup> over NAD<sup>+</sup> as cofactor (FIG. 6) consistent with previously reported plant ADH and PDHs<sup>5,19,23</sup>. In addition to substrate specificity, these three types of plant ADHs and PDHs differ in feedback inhibition by Tyr (FIG. 1D; Table 1)<sup>5</sup>. The canonical AtADH2, was highly sensitive to Tyr ( $IC_{50}$ =38  $\mu$ M), whereas GmPDH1 and MtPDH were insensitive to feedback inhibition by Tyr (up to 8 mM in assays) (FIG. 1D). The noncanonical ADHs, GmncADH and MtncADH, are sensitive to Tyr but with  $IC_{50}$  values in the mM range. Thus, unlike PDHs, legume noncanonical ADHs are partially inhibited by Tyr.

TABLE 1

Steady-state kinetic parameters and effect of tyrosine on representative plant ADH and PDH.					
protein	substrate	$k_{cat}$ ( $s^{-1}$ )	$K_m$ (mM)	$k_{cat}/K_m$ ( $M^{-1} s^{-1}$ )	$IC_{50}^{Tyr}$ (mM)
GmPDH1	prephenate	30.4 ± 0.7	0.09 ± 0.01	337,800	—
	arogenate	6.3 ± 0.7	2.59 ± 0.09	2,430	31.3 ± 9.3
MtPDH1	prephenate	18.5 ± 3.0	0.05 ± 0.01	370,000	—
	arogenate	16.9 ± 1.1	0.94 ± 0.04	17,980	32.1 ± 10.0
Peanut PDH/ADH	prephenate	2.8 ± 0.2	0.19 ± 0.01	14,740	—
	arogenate	3.2 ± 0.1	0.28 ± 0.03	11,430	—
GmncADH	prephenate	—	—	—	—
	arogenate	27.7 ± 1.1	0.41 ± 0.03	67,560	15.5 ± 0.8
MtnADH	prephenate	0.3 ± 0.1	6.69 ± 0.27	45	0.6 ± 0.1
	arogenate	39.0 ± 7.6	0.14 ± 0.02	278,600	2.2 ± 0.5
SolyncADH	prephenate	—	—	—	—
	arogenate	15.9 ± 2.1	0.45 ± 0.05	35,330	12.8 ± 2.0

To further define the phylogenetic boundaries of noncanonical ADH and PDHs additional homologs from *Arachis ipaensis* (peanut; AipaensisVYE8T) and *Solanum lycopersicum* (tomato; Slycopersicum06g050630), which exist at key phylogenetic boundaries (FIG. 1B), were biochemically characterized. AipaensisVYE8T (peanut PDH/ADH) used both arogenate and prephenate to similar degrees ( $k_{cat}/K_m=11.6$  and  $14.9 \text{ mM}^{-1} \text{ s}^{-1}$ , respectively), whereas Slycopersicum06g050630 (SolyncADH) exhibited ADH but not PDH activity (FIG. 7A). Peanut PDH/ADH was insensitive to Tyr inhibition, whereas SolyncADH showed relaxed sensitivity to Tyr with an  $IC_{50}=12.8 \text{ mM}$  (FIG. 7B; Table 1), similar to legume ncADHs. Thus, legume enzymes having considerable PDH activity are Tyr insensitive and form a subclade within the noncanonical clade likely due to a recent gene duplication of an ncADH within legumes (FIG. 1B).

#### X-Ray Crystal Structure of Soybean PDH

To understand the structure-sequence relationship of legume PDHs and ADHs, and because TyrA structures from

plants are not available, the x-ray crystal structure of GmPDH1 was determined by single-wavelength anomalous dispersion phasing using selenomethionine-substituted protein (Table 2). The resulting model was then used for molecular replacement with a 1.69 Å resolution native data set to solve the structure of the GmPDH1•NADP<sup>+</sup>•citrate complex (FIG. 2A; Table 2). GmPDH1 forms a homodimer with each 257 amino acid monomer adopting a N-terminal Rossmann fold domain (residues 8-171) that shapes the NADP(H)-binding domain and an  $\alpha$ -helical C-terminal dimerization domain (residues 172-257) (FIG. 2A). The PDH dimer is formed by two tail-to-tail monomers that pack closely resulting in a dumbbell-shaped molecule (FIG. 2A). The N-terminal domain is made up of seven  $\beta$ -strands sandwiched between two sets of three  $\alpha$ -helices. The C-terminal dimerization domain consists of an entirely helical architecture of four  $\alpha$ -helices. The active site in each monomer is found at the interface of the two domains.

TABLE 2

Summary of crystallographic data.				
Crystal	GmPDH1 (SeMet)•NADP <sup>+</sup>	GmPDH1•NADP <sup>+</sup> •citrate	GmPDH1 N222D•NADP <sup>+</sup> •Tyr	GmPDH1 M219T/N222D•NADP <sup>+</sup> •Tyr
Space group	P1	P1	P1	P1
Cell dimensions	a = 46.51, b = 55.13, c = 68.59 Å; $\alpha = 107.3^\circ$ , $\beta = 98.9^\circ$ , $\gamma = 103.6^\circ$	a = 46.00, b = 55.28, c = 67.94 Å; $\alpha = 107.4^\circ$ , $\beta = 98.9^\circ$ , $\gamma = 103.2^\circ$	a = 46.46, b = 55.05, c = 68.39 Å; $\alpha = 107.8^\circ$ , $\beta = 99.6^\circ$ , $\gamma = 102.6^\circ$	a = 46.29, b = 54.60, c = 68.09 Å; $\alpha = 107.0^\circ$ , $\beta = 99.3^\circ$ , $\gamma = 103.7^\circ$
Date collection				
Wavelength (Å)	0.979	0.979	0.979	0.979
Resolution range (Å) (highest shell)	34.1-2.03 (2.06-2.03)	32.4-1.69 (1.72-1.69)	33.9-1.99 (2.05-1.99)	34.0-1.69 (1.72-1.69)
Reflections (total/unique)	67,565/37,512	126,889/64,687	62,159/36,694	106,188/59,535
Completeness (highest shell)	96.0% (88.7%)	97.0% (94.6%)	88.9% (85.3%)	88.9% (87.2%)
$\langle I/\sigma \rangle$ (highest shell)	7.6 (1.6)	13.8 (1.1)	10.2 (2.3)	12.3 (1.7)
$R_{sym}$ (highest shell)	6.1% (43.2%)	4.4% (43.9%)	8.5% (28.5%)	4.9% (34.4%)
Refinement				
$R_{cryst}/R_{free}$	18.8%/22.9%	15.3%/18.2%	15.8%/20.6%	15.4%/18.4%
No. of protein atoms, waters, ligand atoms	4017, 224, 96	4094, 604, 122	4054, 435, 122	4084, 616, 122
Root mean square deviation, bond lengths (Å)	0.010	0.007	0.008	0.007

TABLE 2-continued

Summary of crystallographic data.				
Crystal	GmPDH1 (SeMet)•NADP <sup>+</sup>	GmPDH1•NADP <sup>+</sup> •citrate	GmPDH1 N222D•NADP <sup>+</sup> •Tyr	GmPDH1 M219T/ N222D•NADP <sup>+</sup> •Tyr
Root mean square deviation, bond angles (°)	1.17	1.17	0.91	0.98
Average B-factor (Å <sup>2</sup> ) protein, ligand, solvent	29.1, 25.3, 37.4	23.7, 19.6, 37.5	32.5, 32.0, 41.3	21.4, 15.3, 37.3
Stereochemistry, most favored, allowed, outliers	97.8, 2.2, 0.0%	97.7, 2.3, 0.0%	97.8, 2.2, 0.0%	97.5, 2.5, 0.0%

15

Consistent with the NADP<sup>+</sup> specificity of GmPDH1<sup>5</sup>, the crystal structure of GmPDH1 shows clear electron density for this ligand in the N-terminal domain of each monomer (FIG. 2B) and extensive protein-ligand binding interactions (FIG. 2C). The (β1a-α1 loop (residues 16-21) is the conserved GxGxxG motif characteristic of NAD(P)(H)-dependent oxidoreductases<sup>7</sup> and contributes interactions with the pyrophosphate moiety and the nicotinamide ring. The main-chain amides of Asn19 and Phe20 hydrogen bond with an oxygen atom in the diphosphate linker. The hydroxyl group of Ser223 interacts with another phosphate oxygen. Additionally, contacts with five water molecules further stabilize the diphosphate linker. The syn-conformation of the nicotinamide ring is stabilized by 7C-7C stacking interactions with Phe20 and by polar contacts between N1 and the side-chain of Ser101. Water molecules also interact with the carboxamide oxygen and nitrogen. These interactions orient the B-face of the nicotinamide ring toward the substrate binding pocket.

Other interactions complete the cofactor binding site (FIG. 2C). The adenine ring, which is in the anti-conformation, hydrogen bonds to the side-chain of Glu80 and a water molecule through its exocyclic N6 and to the hydroxyl group of Thr73 via N3 and N9. Water molecules form polar interactions with the adenine N3 and N7. Extensive charge-charge interactions are formed between the 2'-phosphate of the adenine ribose and the side-chain of Arg40, the hydroxyl groups of Ser39, Ser41, and Tyr43, the backbone amide nitrogen of Ser41, and three water molecules. These interactions form the phosphate binding site that favors NADP (H) over NAD(H). The 3'-phosphate of the adenine ribose interacts with the main-chain amide of Gly18 and the ring oxygen of the ribose hydrogen bonds to the hydroxyl group of Thr73. Both the adenine ribose and the nicotinamide ribose adopt the C2'-endo conformation. The 2'-hydroxyl of the nicotinamide ribose interacts with the side-chain hydroxyl and the main-chain nitrogen of Ser101, whereas the 3'-hydroxyl of the nicotinamide ribose hydrogen bonds to the backbone oxygen of Thr73. A water molecule interacts with the 2'- and 3'-hydroxyls of the nicotinamide ribose.

Although efforts to obtain crystals with different substrate molecules (e.g. prephenate and HPP) were not successful, the structure of PDH complexed with NADP<sup>+</sup> and citrate, contributed from the crystallization buffer, suggests how substrates may bind within the active site (FIG. 2D). The citrate is positioned in a pocket proximal to the nicotinamide ring and the putative catalytic histidine (His124). The Ne of

His124 and the side-chain amine of Gln184 form polar contacts with the α-carboxyl group of citrate. Similarly, the side-chain nitrogen of Gln184 and Ne of His188 contact the γ-hydroxyl of citrate. The ζ-carboxyl group of citrate interacts with the hydroxyl of Thr206, which is provided by the other subunit at the dimer interface. Additional polar contacts are made between the 6-carboxyl and the hydroxyl of Thr131 and the side-chain amine of Asn222. The binding of citrate, which mimics the dicarboxylate portion of prephenate, identifies potential residues in the substrate binding site. Identification of a Residue that Confers TyrA Substrate Specificity

Next, the predicted substrate binding site (FIG. 2D) and the phylogenetic distribution of PDH and ADHs (FIG. 1B) were used together to identify residues responsible for differences in substrate specificity. Amino acid alignment of the plant TyrA enzymes (FIG. 8) showed highly conserved residues responsible for NADP<sup>+</sup> binding, including the GxGxxG motif, and residues proposed to function in catalysis (e.g. Ser101 and His124)<sup>31-34</sup>. Within the PDH active site, residues uniquely conserved in either ADHs or PDHs were also identified (FIG. 3A; FIG. 8). Asp218 in GmncADH, which corresponds to Asn222 in GmPDH1, was highly conserved among ADHs but not in PDHs (FIG. 3A). Similarly, Thr215 of GmncADH was generally conserved among ADHs but replaced by either Met or Val in PDHs (Met219 in GmPDH1); however, peanut PDH/ADH retains a Thr at the corresponding position (FIG. 3A). These comparisons suggest that either Met219 or Asn222 (or both) may determine prephenate specificity in PDH.

To experimentally test the roles of the two residues in PDH versus ADH substrate specificity, site-directed mutagenesis was performed on GmPDH1 to convert Asn222 and Met219 into the corresponding residues in GmncADH (N222D and M219T). The M219T mutant had very similar kinetic parameters to wild-type enzyme preferring prephenate over arogenate substrate (FIG. 3B; Table 3). The N222D mutant, however, showed a 115-fold reduction in  $k_{cat}/K_m$  with prephenate and gained ADH activity (FIG. 3B; Table 3). The turnover rate ( $k_{cat}$ ) of N222D for arogenate (27.8 s<sup>-1</sup>) was comparable to wild-type GmPDH1 and GmncADH for prephenate and arogenate, respectively (30.4 and 27.7 s<sup>-1</sup>; Table 1). The M219T/N222D double mutant, exhibited very similar  $k_{cat}/K_m$  values for PDH and ADH activity compared to the N222D single mutant (FIG. 3B; Table 3), suggesting that the M219T substitution had little effect on substrate specificity alone or in combination with the N222D mutation.

TABLE 3

Steady-state kinetic parameters and effect of tyrosine on mutant GmPDH1, MtPDH1, GmncADH, MtnCADH, and SolyncADH.					
protein	substrate	$k_{cat}$ ( $s^{-1}$ )	$K_m$ (mM)	$k_{cat}/K_m$ ( $M^{-1} s^{-1}$ )	$IC_{50}^{Tyr}$ (mM)
GmPDH1 M219T	prephenate	$30.3 \pm 0.8$	$0.10 \pm 0.02$	303,000	—
	arogenate	$4.6 \pm 0.3$	$1.55 \pm 0.14$	2,968	—
GmPDH1 N222D	prephenate	$19.1 \pm 1.4$	$7.58 \pm 1.13$	2,520	$5.3 \pm 0.4$
	arogenate	$27.8 \pm 3.5$	$0.53 \pm 0.18$	52,450	$4.7 \pm 0.4$
GmPDH1 N222A	prephenate	$6.6 \pm 0.3$	$0.19 \pm 0.04$	34,740	—
	arogenate	—	—	—	—
GmPDH1 M219T/N222D	prephenate	$2.5 \pm 0.3$	$1.18 \pm 0.12$	2,119	$11.1 \pm 1.2$
	arogenate	$29.0 \pm 4.4$	$0.63 \pm 0.18$	46,030	$5.9 \pm 0.5$
MtPDH C220D	prephenate	$0.5 \pm 0.1$	$1.53 \pm 0.09$	327	$8.2 \pm 0.9$
	arogenate	$46.8 \pm 4.2$	$0.27 \pm 0.01$	173,300	—
GmncADH D218N	prephenate	$11.5 \pm 0.5$	$1.98 \pm 0.05$	5,810	—
	arogenate	$8.6 \pm 0.2$	$0.74 \pm 0.14$	11,620	—
MtnCADH D220C	prephenate	$10.1 \pm 0.5$	$0.74 \pm 0.03$	13,650	—
	arogenate	$7.0 \pm 0.8$	$0.87 \pm 0.03$	8,046	$7.7 \pm 1.5$
SolyncADH D224N	prephenate	$2.4 \pm 1.0$	$2.31 \pm 0.10$	1,040	—
	arogenate	$11.7 \pm 0.2$	$1.34 \pm 0.48$	8,730	—

To test if the analogous mutation alters substrate specificity outside of soybean PDH, the Asp residue was introduced to the corresponding Cys on MtPDH. Similar to the GmPDH1 N222D mutant, the C220D mutation reduced PDH activity and enhanced ADH activity (FIG. 3B), which is reflected by 31-fold higher and 3-fold lower  $K_m$  toward prephenate and aroenate, respectively, compared to wild-type (Tables 1 and 3). To examine if an acidic Asp residue was necessary for converting PDH to ADH activity, an alanine mutation was introduced at Asn222 in GmPDH1 (N222A). The N222A mutant reduced PDH activity, but did not introduce ADH activity, unlike N222D (FIG. 3B; Table 3). These results suggest that the corresponding 222 position in legume PDH enzymes is the key determinant for their substrate specificity, where an acidic Asp residue is crucial for ADH activity.

#### Altered Substrate Specificity Simultaneously Affects Tyr-Sensitivity

The mutations on legume PDHs were also tested for their effect on Tyr sensitivity. Similar to GmPDH1, the M219T and N222A single mutants, which did not alter substrate specificity, were not inhibited by Tyr (FIG. 3C; Table 3). In contrast, the GmPDH1 N222D and M219T/N222D mutants, as well as the MtPDH C220D mutant, exhibited Tyr inhibition with  $IC_{50}$  values of 5 to 11 mM (FIG. 3C; Table 3). Thus, mutating Asn222 and Cys220 of GmPDH1 and MtPDH, respectively, into an Asp not only introduced ADH activity, but also Tyr sensitivity.

The GmPDH1 mutants that bind to Tyr can now be used to test the role of the active site Asp222 in ADH activity and Tyr sensitivity. The GmPDH1 N222D and M219T/N222D mutants were successfully co-crystallized with Tyr and NADP<sup>+</sup> bound in their active site at 1.99 and 1.69 Å resolution, respectively (Table 2). An overlay of these two mutants with the wild-type structure revealed no global conformational changes (FIG. 4A). Likewise, the substitutions did not drastically alter the active site structure of either mutant (FIG. 4B).

In the GmPDH1 M219T/N222D structure, the ring hydroxyl of the Tyr ligand contacts NE of His124, the hydroxyl of Ser101, and the amine group of Gln184 (FIG. 4C). The side chain carboxylate of Tyr interacts with the hydroxyl group and backbone amide of Thr131, as well as the carbonyl and backbone amide of Gln130. The position of the bound Tyr is also stabilized by  $\pi$ - $\pi$  stacking interaction with the nicotinamide ring of NADP<sup>+</sup>. The amine nitrogen

of Tyr forms polar contacts with a water molecule, the carbonyl of Gln130, and the carboxylate of the mutated Asp222 residue. Identical contacts were observed in the GmPDH1 N222D structure. Neither Met219 nor the mutated Thr219 makes a direct contact with the ligand.

In the GmPDH1 mutant structures, the active site pocket near the site of hydride transfer from the substrate to the nicotinamide via His124 is composed of a wall of nitrogen atoms (i.e. of Gln184 and His188), and Asp222 adds a negatively charged region to the side of the pocket to recognize the amine of Tyr (FIG. 4C). Computational docking of aroenate into the crystallographic structure of GmPDH1 M219T/N222D shows that the hydroxyl of aroenate can anchor itself between His124 and the nicotinamide ring, similar to Tyr (FIG. 4D). Also, the carboxylate of Asp222 forms a polar interaction with the amine of aroenate (FIG. 4D). By mutating the 222 residue from a positively charged Asn to a negatively charged Asp, the specificity in substrate recognition changes to preferentially recognize the amine of aroenate over the carbonyl of prephenate and also introduce sensitivity to Tyr.

#### Mutating Asp218 Introduces PDH Activity in Divergent Plant ADH Enzymes

To test if PDH activity can be introduced to legume ncADHs, the reciprocal mutation was made on GmncADH at position Asp218 (corresponding to Asn222 of GmPDH1) to generate the D218N mutant. The D218N substitution reduced  $k_{cat}/K_m$  for ADH by ~6-fold (FIG. 5A; Table 3) while introducing PDH activity (FIG. 5A) into an enzyme which was originally unable to use prephenate (FIG. 1C; Table 1). The corresponding Asp to Cys mutation on MtnCADH (D220C) showed similar results, e.g. reduced ADH activity and enhanced PDH activity (FIG. 5A). While wild-type MtnCADH had a 6,190-fold preference for aroenate, MtnCADH D220C was switched to prefer prephenate over aroenate by 1.7-fold (Table 3). These results further confirm the role of Asp218 and Asn222 in ADH and PDH activity, respectively.

The corresponding Asp residue was also mutated to Asn in divergent ADH from the basal noncanonical clade, tomato (SolyncADH D224N), and canonical ADH clade, Arabidopsis (AtADH2 D241N) (FIG. 1B). Similar to the results observed with the legume ncADHs, the tomato and Arabidopsis mutant enzymes gained PDH activity at the expense of ADH activity (FIG. 5A; Table 3). Additionally, each of the ADH mutants (GmncADH D218N, MtnCADH D220C,

SolyncADH D224N, and AtADH2 D241N) were significantly less sensitive to Tyr inhibition than the respective wild-type enzymes (FIG. 5B; Tables 1 and 3). Thus, the alteration of the key active site Asp residue is the evolutionary switch needed to introduce PDH activity in diverse plant ADH enzymes while simultaneously relieving feedback inhibition by Tyr.

### Discussion

In plants, aromatic amino acid biosynthesis provides essential building blocks for proteins and diverse primary and specialized metabolites<sup>6,7</sup>; however, the biochemical pathways for production of these compounds can vary, as exemplified in Tyr biosynthesis. While all plants have canonical ADH for Tyr synthesis<sup>5-7,19,37</sup>, our studies found that some eudicots have noncanonical ADH (ncADH) and some legumes additionally have PDH (FIG. 1B-1C)<sup>5</sup>. The three types of TyrA dehydrogenases share similar catalytic properties, but with distinct arogenate versus prephenate specificities (FIG. 1C; Table 1; FIGS. 7 & 8)<sup>5,19,24,27,28</sup>. Moreover, the final pathway product, Tyr, strongly feedback inhibits the canonical ADHs and partially inhibits the ncADHs (FIG. 1D), whereas the legume PDHs are completely insensitive to Tyr (FIG. 1D)<sup>5</sup>. Also, unlike plastid-localized canonical ADHs<sup>19,37,38</sup>, ncADH and PDH lack an N-terminal chloroplast transit peptide and localize in the cytosol<sup>5</sup>, as were also shown for cytosolic CM and TAT isoforms that function before and after PDH, respectively<sup>39,40</sup>. While we are currently investigating the physiological functions of the cytosolic PDH and ADH pathways using genetic approaches, our data suggest that alternative Tyr pathways having distinct regulation and localization evolved in different plants.

Previous work showed that the legume PDH genes evolved through duplication of an ancestral plant ADH gene, followed by subfunctionalization, rather than horizontal gene transfer of a bacterial PDH gene<sup>5</sup>. PDH enzymes are restricted to legumes, particularly in the more recently-diverged species, such as peanut and soybean (FIG. 1B; FIG. 9)<sup>41,42</sup>. Therefore, the PDH genes evolved through an ancient duplication event giving rise to the eudicot noncanonical clade, which was followed by a second duplication within the legume family (FIG. 1B, FIG. 9).

The current study demonstrates that alteration of Asp222 (into Asn or Cys) played a key role during the subfunctionalization of the duplicated gene from ADH to PDH (FIGS. 3 and 5). Comparison of the x-ray crystal structures of the wild type and N222D mutants of GmPDH1 (FIGS. 2 and 4) showed that the Asp substitution is readily accommodated in

the active site without significant conformational changes (FIGS. 4A, 4B). Prephenate and arogenate are nearly identical with the exception of a carbonyl versus an amine, respectively (FIG. 1A). Positioning of the carboxylate side-chain of the Asp residue in the GmPDH1 mutants provides an energetically dominant ionic interaction with the amine of arogenate substrate (FIG. 4D), which would be protonated at physiological pH, compared to a hydrogen bond with the prephenate carbonyl group in the wild-type enzyme. The same charge-charge interaction is also critical for feedback inhibition in the GmPDH1 mutants (FIG. 3C; Table 3) and binding with Tyr, which also has the side-chain amine (FIG. 4C).

Although introduction of Asp218 into GmPDH1 restored ADH activity near wild-type levels of GmncADH ( $k_{cat}/K_m$  of 52.5 vs 67.5, respectively FIG. 3B, Tables 1 and 3), that of Asn222 into GmncADH was insufficient to obtain PDH activity comparable to wild-type GmPDH1 level (FIGS. 3B and 5A). An additional mutation of Met219, which covaries with Asn222, on GmPDH1 wild-type and N222D mutant did not enhance ADH activity (FIG. 3B). Comparisons among GmPDH1, GmncADH, and AtADH2 reveal variety in the amino acid sequence of the 131e-131f loop (Phe127 to Trp136 in GmPDH1, FIG. 8), which is at the opposing side of the active site from Asn222 and consists of residues that interact with the ligand side chain carboxylate (FIG. 4B). Thus, residues on the  $\beta$ 1e- $\beta$ 1f loop could be contributing to the correct positioning of the substrate for catalysis, and various combinations of active site mutations at both sides may be needed to convert an ADH to a fully functional PDH.

The residue corresponding to Asp218 that confers ADH activity can now be used to trace the evolutionary origin of the plant ADHs. Asp218 is present in TyrA homologs of all plants and algae, including green, red, and brown algae (FIG. 10), suggesting that Asp218-containing ADH enzymes are universal to the plant kingdom. Previous and current analyses showed that plant ADHs are more closely related to proteobacteria and methanogens (archaea) than cyanobacteria<sup>16,21,43</sup> (FIG. 10B, Table 4). Interestingly, an Asp residue was present at the corresponding 218 position in the TyrA orthologs of proteobacteria, which was previously shown to have ADH activity (e.g. *Phenylobacterium immobile*<sup>26</sup>), but absent in those of archaea (FIGS. 10B, 10C). Together these data suggest that ADH enzymes containing Asp218 evolved in a bacteria ancestor, which was horizontally transferred to the common ancestor of plants and algae. Together with PPA-ATs acquired from a Chlorobi/Bacteroidetes ancestor, the Asp218-containing ADHs are maintained in the plant kingdom to synthesis Tyr via the arogenate pathway.

TABLE 4

Amino acid sequence similarity comparison for representative plant and microbial TyrA homologs.									
	GmPDH1	GmncADH	AtADH2	<i>D. multivorans</i>	<i>T. xiamensis</i>	<i>M. harundinacea</i>	<i>Synechocystis</i> ADH	<i>A. aeolicus</i> PDH	<i>S. cerevisiae</i> PDH
GmPDH1	100								
GmncADH	90.0	100							
AtADH2	68.75	69.37	100						
<i>D. multivorans</i>	62.08	63.54	54.37	100					
<i>T. xiamensis</i>	61.87	61.87	55.0	66.45	100				
<i>M. harundinacea</i>	53.75	53.33	52.7	53.29	55.41	100			
<i>Synechocystis</i> ADH	52.33	53.0	47.19	56.04	53.16	47.08	100		

TABLE 4-continued

Amino acid sequence similarity comparison for representative plant and microbial TyrA homologs.									
	GmPDH1	GmncADH	AtADH2	<i>D.</i> <i>multivorans</i>	<i>T.</i> <i>xiamensis</i>	<i>M.</i> <i>harundinacea</i>	<i>Synechocystis</i> ADH	<i>A.</i> <i>aeolicus</i> PDH	<i>S.</i> <i>cerevisiae</i> PDH
<i>A. aeolicus</i> PDH	47.5	48.54	42.91	47.91	48.75	42.5	56.25	100	
<i>S. cerevisiae</i> PDH	21.66	20.83	21.66	18.54	22.08	25.41	17.7	12.7	100

Sequence similarity is based on the network shown in FIG. 10. *S. cerevisiae* PDH was included because it was found to be sister to plants in some phylogenetic analyses<sup>16,30</sup>; however, due to lack of sequence similarity it was not present in our sequence similarity network or other phylogenetic analyses of plant TyrA homologs<sup>1,43</sup>. Plant TyrA homologs share greater sequence similarity with proteobacterial TyrA homologs than archaea, cyanobacteria, yeast, or other bacteria.

Is the corresponding Asp residue also responsible for substrate specificity and regulation of divergent microbial TyrA dehydrogenases? To address this question, the three-dimensional structure of GmPDH1 (FIG. 2), the first of a plant TyrA structure, was compared to previously reported microbial TyrAs from the cyanobacteria *Synechocystis* sp. PCC 6803 (SynADH; PDB: 2F1K;<sup>29</sup>) and *A. aeolicus* PDH (AaPDH; PDB: 3GGG;<sup>34</sup>). SynADH is specific to arogenate substrate and Tyr insensitive, whereas AaPDH prefers prephenate and is sensitive to Tyr<sup>29,34</sup>. The overall fold of GmPDH1 is conserved (root mean square deviations of 2.5-3.0 Å<sup>2</sup> for ~235 Cα atoms) with SynADH and AaPDH (FIG. 11A). While the N-terminal Rossmann-fold was highly conserved, some differences in topology were found in the C-terminal dimerization domain: the 3<sub>10</sub> helix (α9) and the long C-terminal helix (α13) of SynADH and AaPDH are missing in the soybean enzyme, and the α7 helix of GmPDH1 is split into two helices in SynADH and AaPDH (α7 and α8) (FIG. 11A).

Comparison of cofactor binding sites reveals a structural variation near the adenine ribose, which defines NADP(H) cofactor specificity of GmPDH1. An elongated β1b-α2 loop in GmPDH1 (Ser39-Tyr43) and also NADP(H)-dependent SynADH (Ser30-Thr35) forms charge-charge and hydrogen bond contacts with the phosphate group of NADP(H). In contrast, the shorter loop of NAD(H)-dependent AaPDH (Asp62-Ile63) fills the corresponding space and allows for direct interaction with the hydroxyl groups of the adenine ribose of NAD(H) (FIG. 11). Interestingly, the diphosphate group of NADP(H) adopts a trans-conformation in GmPDH1, where the same cofactor moiety in SynADH and AaPDH are in cis-conformations (FIG. 11B). In SynADH and AaPDH, a 4.5 and 7.7 Å shift in α1 compared to GmPDH, respectively, containing part of the GxGxxG motif, accommodates the cis conformer of cofactor. Thus, the trans-conformation of cofactor appear to be a unique feature of GmPDH1 and likely plant TyrAs.

Despite the cofactor binding site variations, each structure maintains the positioning of the ribose and nicotinamide ring relative to a key catalytic histidine (FIGS. 4B, 4C; FIGS. 11B, 11C). The residues that contribute hydrogen bonds to the nicotinamide ribose (Thr73 and Ser101 in GmPDH1; Thr65 and Ser92 in SynADH; Ser99 and Ser126 in AaPDH) are conserved, as is an apolar residue stacking with the nicotinamide ring (Phe28 in GmPDH1; Ile11 in SynADH; Met41 in AaADH) (FIGS. 4B, 4C; FIGS. 11B, 11C). Overall, these interactions position the C4 of the nicotinamide ring in proximity to the conserved catalytic histidine (His124 in GmPDH1; His112 in SynADH; His147 in AaPDH) for the ensuing oxidative decarboxylation reaction<sup>29,33,34</sup>.

Notable differences were found in the architecture of the residues and regions that recognize the side chain of sub-

strates and the Tyr effector (FIG. 11C); part of which reflects the structural variations in the dimerization domain (FIG. 11A). SynADH contains an Asn in the 222 position similar to GmPDH1, while AaPDH has Asp255 at the corresponding position. However, the placement of α-helix adjacent to Asn222 or Asp255 (all in SynADH and AaPDH compared to α9 in GmPDH1) varies. This is likely due to a proline residue uniquely present in SynADH and AaPDH but absent in GmPDH1, which kinks the all helix to orient the ligand towards the catalytic His. Moreover, the 131e-131f loop, which is opposite from Asn222 or Asp255, is condensed in GmPDH1 (Phe127-Trp136) compared to SynADH (Ala115-Leu129) and AaPDH (Ala150-Leu164). These key differences in the active site configuration likely prevent the Asp/Asn residue from being involved in arogenate/prephenate specificity and Tyr inhibition in the microbial structures (FIG. 11C). Thus, microbial TyrA dehydrogenases, which are distantly-related from plant TyrAs (Table 4), have taken different and yet unknown evolutionary pathway towards refining substrate specificity as compared to plant TyrAs.

In summary, using a combined phylogenic and structural approach, we identified the critical residue that controls the substrate specificity and Tyr sensitivity of TyrAs and underlies the functional evolution of alternative Tyr pathways in plants. The high conservation of the Asp residue among all plantae and some microbial TyrA orthologs suggests an ancient evolutionary origin of the ADH Tyr pathway universally present in the plant kingdom today. The identified key residue can now be used to alter Tyr biosynthetic pathways and regulation, as demonstrated in diverse plant TyrAs (FIG. 5), to optimize Tyr availability for the production of its derived natural products, including vitamin E and morphine alkaloid.

#### Generation of Transgenic Plants

The ADH and PDH polynucleotides, constructs and vectors described herein may be used to generate transgenic plants comprising the ADH and PDH polynucleotides. The ADH and PDH polynucleotides will be operably connected to a promoter functional in the plant cells. The resulting construct will be introduced into the plant cells via a method of transformation or other introduction of genetic material into plant cells. One optional method is insertion via *Agrobacterium tumefaciens* insertion of the DNA into the flowering plants. The polynucleotide can then be selected for either directly by testing for expression of the inserted polynucleotide or alternatively the construct may include a selectable marker to make selection of transgenic plants simple.

#### Materials and Methods

##### Identification of ncADH Enzymes from Plants

BlastP® alignment tool searches were performed using the amino acid sequence of GmPDH1/Gm18g02650 (KM507071) and MtPDH/Mt3g071980 (KM507076) as

queries against various plant lineages found within the Phytozome (phytozome.net) and 1KP (onekp.com) databases. A phylogenetic analysis was performed using all the homologs identified through BlastP® alignment tool searches. Evolutionary distances were estimated based on maximum likelihood<sup>44</sup>. Phylogenetic analysis was performed in MEGA6<sup>45</sup> from an amino acid alignment using MUSCLE<sup>46</sup>. All positions with <75% site coverage were removed, leaving 263 positions in the final analysis from 32 sequences, the tree was estimated with 1,000 bootstrap replicates (FIG. 1B).

Recombinant Protein Expression and Purification and Site Directed Mutagenesis

Full-length coding sequences of GmPDH1, GmncADH, MtPDH, MtnADH were amplified using gene-specific primers with Phusion DNA polymerase (Thermo). The PCR

products were purified using QIAquick gel extraction kit (Qiagen) and ligated into pET28a vector (Novagen) at EcoRI and NdeI sites, in frame with an N-terminal 6x-His tag using In-Fusion HD cloning kit and protocol (Clontech). A PCR reaction consisting of 1 U Phusion DNA polymerase (Thermo) with 0.2 mM dNTP's, 0.5 μM forward and reverse primers (Table 5) and 1x Phusion reaction buffer (Thermo) were mixed with plasmid template diluted 100-fold. The mixture was placed in a thermocycler for 98° C. for 30 s followed by 20 cycles of 10 s at 98° C., 20 s at 70° C., 4.5 min at 72° C. with a final extension at 72° C. for 10 min. PCR products were purified using a QIAquick Gel Extraction Kit, then treated with DpnI (Thermo) to digest methylated template DNA for 30 min at 37° C. Plasmids encoding either wild-type or site-directed GmPDH1 were transformed into *E. coli* XL1-Blue cells, and sequenced to confirm the correct mutation was made.

TABLE 5

Primers used in this Example		
Name	Use	sequence (5'-3')
GmPDHM219TF	mutagenesis	GGAGACGACGATGAGAAATAGTTTGTATTGTATAG (SEQ ID NO: 97)
GmPDHM219TR	mutagenesis	CAAACTATTTCTCATCGTCCTTCAATTTAAC (SEQ ID NO: 98)
GmPDHN222DF	mutagenesis	GGAGACGATGATGAGAGATAGTTTGTATTGTATAG (SEQ ID NO: 99)
GmPDHN222DR	mutagenesis	CAAACTATCTCTCATCGTCCTTCAATTTAAC (SEQ ID NO: 100)
GmPDHN222AF	mutagenesis	GGAGACGATGATGAGAGCTAGTTTGTATTGTATAG (SEQ ID NO: 101)
GmPDHN222AR	mutagenesis	CAAACTAGCTCTCATCGTCCTTCAATTTAAC (SEQ ID NO: 102)
GmPDHM219TN222DF	mutagenesis	GGAGACGACGATGAGAGATAGTTTGTATTGTATAG (SEQ ID NO: 103)
GmPDHM219TN222DR	mutagenesis	CAAACTATCTCTCATCGTCCTTCAATTTAAC (SEQ ID NO: 104)
GmncADHD218NF	mutagenesis	AGGACACCACCATCAGAAATAGTTTGTACTGTACA (SEQ ID NO: 105)
GmncADHD218NR	mutagenesis	AAACTATTTCTGATGGTGTCTTCAATTGAA (SEQ ID NO: 106)
MtPDHC220DF	mutagenesis	GTCATGGGTGATAGTTTGTATAGTGGATTATTCG (SEQ ID NO: 107)
MtPDHC220DR	mutagenesis	GATCAAACTATCACCATGACAGTTTTTCAACTCAAC (SEQ ID NO: 108)
MtnADHF	cloning	CGCGCGCAGCCATATGTCAAATTCACCTTCTCTG (SEQ ID NO: 109)
MtnADHR	cloning	GACGGAGCTCGAATTCATGCATCAACATTCAGTCTT (SEQ ID NO: 110)
MtnADHD220CF	mutagenesis	CCATGAGATGTAGTTTGTATCTGTACAGTGGATTGTTG (SEQ ID NO: 111)
MtnADHD220C	mutagenesis	CAAACTACATCTCATGGTGGTCTTTCAGTTGAGTAAG (SEQ ID NO: 112)
Peanut PDH/ADHF	cloning	CGCGCGCAGCCATATGTCACTTCCCATCCCCAAA (SEQ ID NO: 113)
Peanut PDH/ADHR	cloning	GACGGAGCTCGAATTCACCTTTCAGTTTTTCTT (SEQ ID NO: 114)

TABLE 5-continued

Primers used in this Example		
Name	Use	sequence (5'-3')
SolyncADHF	cloning	CGCGCGGCAGCCATATGATGTCTTCATCTTCTTCTTG (SEQ ID NO: 115)
SolyncADHR	cloning	GACGGAGCTCGAATCTTAGAACTTTGATATGATAGG (SEQ ID NO: 116)
SolyncADHD224NF	mutagenesis	GCTCAGTTAAAAATAGTTTTGATCTGTTTCAGCGG (SEQ ID NO: 117)
SolyncADHD224NR	mutagenesis	GATCAAACTATTTTTAACTGAGCTCTCCTTCAC (SEQ ID NO: 118)
AtADH2D241NF	mutagenesis	CACATCGAGTAATAGCTTTGAGCTTTTCTACGG (SEQ ID NO: 119)
AtADH2D241NR	mutagenesis	CTCAAAGCTATTACTCGATGTGTTCCACCAAATC (SEQ ID NO: 120)

Confirmed plasmids were then transformed into *E. coli* Rosetta-2 (DE3) cells (Novagen) by heat shock at 42° C. for 60 s. For recombinant protein expression, overnight cultures in 10 mL Luria broth (LB) supplemented with 100 µg/mL kanamycin were grown at 37° C. with 200 r.p.m. shaking. The following morning 1 mL of culture was added into 50 mL of fresh LB without antibiotics and allowed to grow at 37° C. with 200 r.p.m. shaking. After 1 hour, 10 mL was added into 500 mL of fresh LB with kanamycin (100 µg/mL) and grown until the OD<sub>600</sub> reached 0.3, and the incubator was changed to 18° C. After 1 hour isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.4 mM final concentration) was added to induce recombinant protein expression and grown for an additional 20 hours. Cultures were spun at 10,000×g for 10 minutes, and the supernatant was decanted. The pellet was resuspended in 100 mL of 0.9 M NaCl, and spun for 10 minutes at 10,000×g. The supernatant was decanted and the remaining pellet was redissolved in 25 mL lysis buffer (25 mM HEPES pH 7.6, 50 mM NaCl, 10% (v/v) ethylene glycol) plus 0.5 mM phenylmethylsulfonyl fluoride. Cells were frozen in liquid N<sub>2</sub>, and thawed in hot water to initiate cell lysis, 25 mg of lysozyme (Dot Scientific) was added and cells sonicated for 3 min. Cell debris was pelleted by centrifugation (30 min; 50,000×g). Supernatant was applied to a 1 mL HisTrap FF column for purification of the His-tagged recombinant protein using an ÄKTA FPLC system (GE Healthcare Bio-Sciences). After loading protein the column was washed with 90% buffer A (0.5 M NaCl, 0.2 M NaP and 20 mM imidazole) and 10% buffer B (0.5 M NaCl, 0.2 M NaP and 0.5 M imidazole, recombinant enzyme was then eluted with 100% buffer B. Fractions containing purified protein were pooled and desalted by Sephadex G50 column (GE Healthcare) size-exclusion chromatography into lysis buffer. The purified proteins were analyzed by SDS-PAGE to determine purity. All protein purification steps were performed at 4° C. unless stated otherwise.

GmPDH1 Crystallization

Purified protein (see above) was loaded onto a Superdex-75 26/60 HiLoad FPLC size-exclusion column (GE Healthcare) equilibrated with 25 mM Hepes, pH 7.5, and 100 mM NaCl. Protein concentration was determined by the Bradford method (Protein Assay, Bio-Rad) with bovine serum albumin as a standard. For selenium-methionine (SeMet) GmPDH1 expression, *E. coli* Rosetta II (DE3) cells containing the PDH construct were grown to an OD<sub>600</sub>~0.6 in M9 minimal media, at which point the media was supple-

mented with 60 mg SeMet, valine, leucine, and isoleucine and 100 mg of lysine, phenylalanine, and threonine and induced with 1 mM IPTG for 16-18 hours at 16° C. SeMet GmPDH1 was purified as described for native GmPDH1.

Purified enzyme was concentrated to 10 mg ml<sup>-1</sup> and crystallized using the hanging-drop vapor-diffusion method with a 2-µl drop (1:1 protein and crystallization buffer). Tyr (3 mM final) was added to both GmPDH1 M219T/N222D and GmPDH1 N222D. Diffraction quality crystals of the native GmPDH1 were obtained at 4° C. with a crystallization buffer of 20% PEG-4000, 30% (w/v) D-sorbitol, and 100 mM sodium citrate, pH 5.5. Crystals of SeMet PDH1 formed at 4° C. with a crystallization buffer of 20% (w/v) PEG-3350, 100 mM sodium citrate, pH 4.0, and 200 mM sodium citrate tribasic. Crystals of GmPDH1 N222D formed in 2 mM of an oxometalates solution containing 0.005 M sodium chromate tetrahydrate, 0.005 M sodium molybdate dihydrate, 0.005 M sodium tungstate dihydrate, and 0.005 M sodium orthovanadate, 0.1 M of MOPSO and bis-Tris, pH 6.5, and 50% (v/v) of a precipitant mixture of 20% (w/v) PEG-8000 and 40% (v/v) 1,5-pentanediol<sup>47</sup>. Crystals of GmPDH1 M219T/N222D formed in 16% (w/v) PEG 8000, 40 mM potassium phosphate dibasic, and 20% (v/v) glycerol. All crystals were flash-frozen in liquid nitrogen with mother liquor supplemented with 25% glycerol as a cryoprotectant.

The GmPDH1 structure was solved by single-wavelength anomalous dispersion (SAD) phasing. Diffraction data collected at beamline 19ID of the Argonne National Laboratory Advanced Photon Source were indexed, integrated, and scaled using HKL3000<sup>48</sup>. SHELX<sup>49</sup> was used to determine initial SeMet positions and to estimate initial phases from the peak wavelength data set. SeMet positions and parameters were refined in MLPHARE<sup>50</sup>. Solvent flattening was performed with DM<sup>51</sup>, and ARP/wARP<sup>52</sup> was used to build an initial model. Iterative rounds of manual model building and refinement were performed with COOT<sup>53</sup> and PHENIX<sup>54</sup>, respectively. The resulting model was used for molecular replacement into the higher resolution native data set using PHASER<sup>55</sup>. Iterative rounds of manual model building and refinement, which included translation-libration-screen (TLS) models, used COOT and PHENIX, respectively. The native GmPDH1 structure was used for molecular replacement to solve the GmPDH1 N222D and GmPDH1 M219T/N222D structures. Each mutant structure was built and refined using the same method as the wild-type

enzyme. Data collection and refinement data are summarized in Table 2. The final model of SeMet-substituted GmPDH1 included residues Ser9 to Gln258 and NADP<sup>+</sup> for both molecules in the asymmetric unit and 228 waters. The final model of the GmPDH1•NADP<sup>+</sup>•citrate complex included residues Gln8 to Ile257 for chain A and residues Gln8 to Thr260 for chain B, NADP<sup>+</sup> and citrate in both chains, and 605 waters. The structure was intended to be an apoenzyme, but NADP<sup>+</sup> and citrate were bound in the active site. The final model of the GmPDH1 N222D•NADP<sup>+</sup>•Tyr complex included residues Ser9 to Met258 for chain A and residues Gln8 to Thr260 for chain B, NADP<sup>+</sup> and Tyr in both chains, and 435 waters. The final model of the GmPDH1 M219T/N222D•NADP<sup>+</sup>•Tyr complex included residues Ser9 to Ile257 for chain A and residues Gln8 to Ile257 for chain B, NADP<sup>+</sup> and Tyr in both chains, and 616 waters. ADH and PDH Assay

Kinetic parameters of purified recombinant proteins were determined from assays conducted at varying aroenate (19.5 μM-5 mM) and prephenate concentration (23.4 μM-6 mM). Standard assay conditions were 25 mM HEPES pH 7.6, 50 mM KCl and 10% (v/v) ethylene glycol, and 0.5 mM NADP<sup>+</sup> with varied substrate, concentrations. Reactions were initiated by addition of enzyme and incubated at 37° C. monitored every 10-15 seconds at A<sub>340nm</sub> using a microplate reader (Tecan Genios). Kinetic parameters were determined by fitting initial velocity data to the Michaelis-Menten equation using the Origin software (OriginLab). Aroenate was prepared by enzymatic conversion of prephenate (Sigma-Aldrich) as previously reported<sup>56</sup>. For Tyr inhibition assays, Tyr was dissolved in a slightly basic solution (0.025 N NaOH) due to solubility issues, thus the concentration of lysis buffer was increased to 500 mM HEPES final concentration to buffer against the changes by addition of Tyr in the reaction. Reactions containing varying amounts of Tyr (10 μM-8 mM) with 0.5 mM NADP<sup>+</sup> and either 1 mM aroenate or 0.8 mM prephenate were monitored as above.

#### Computational Substrate Docking

Molecular docking of prephenate and aroenate into the GmPDH1 M219T/N222D•NADP<sup>+</sup>•Tyr three-dimensional model with Tyr removed was performed using AutoDock Vina (ver. 1.1.2)<sup>57</sup>. The positions of NADP<sup>+</sup> and Tyr in the structure was used to guide docking with a grid box of 30x30x30 Å and the level of exhaustiveness set to 8.

#### REFERENCES FOR EXAMPLE 1

- Weng, J.-K., Philippe, R. N. & Noel, J. P. The Rise of Chemodiversity in Plants. *Science* 336, 1667-1670 (2012).
- Gowik, U. & Westhoff, P. The Path from C3 to C4 Photosynthesis. *Plant Physiol.* 155, 56-63 (2011).
- Torruella, G., Suga, H., Riutort, M., Peretó, J. & Ruiz-Trillo, I. The evolutionary history of lysine biosynthesis pathways within eukaryotes. *J. Mol. Evol.* 69, 240-248 (2009).
- Jensen, R. A. & Pierson, D. L. Evolutionary implications of different types of microbial enzymology for L-tyrosine biosynthesis. *Nature* 254, 667-671 (1975).
- Schenck, C. A., Chen, S., Siehl, D. L. & Maeda, H. A. Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nat. Chem. Biol.* 11, 52-57 (2015).
- Maeda, H. & Dudareva, N. The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.* 63, 73-105 (2012).

- Tzin, V. & Galili, G. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. *Mol. Plant* 3, 956-72 (2010).
- Fernstrom, J. D. & Fernstrom, M. H. Tyrosine, phenylalanine, and catecholamine synthesis and function in the brain. *J. Nutr.* 137, 1539S-1547S; discussion 1548S (2007).
- Hagel, J. M. & Facchini, P. J. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647-672 (2013).
- Petersen, M. et al. Evolution of rosmarinic acid biosynthesis. *Phytochemistry* 70, 1663-1679 (2009).
- Gleadow, R. M. & Møller, B. L. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu. Rev. Plant Biol.* 65, 155-185 (2014).
- Strack, D., Vogt, T. & Schliemann, W. Recent advances in betalain research. *Phytochemistry* 62, 247-69 (2003).
- Hunter, S. C. & Cahoon, E. B. Enhancing vitamin E in oilseeds: unraveling tocopherol and tocotrienol biosynthesis. *Lipids* 42, 97-108 (2007).
- Millner, P. a. & Barber, J. Plastoquinone as a mobile redox carrier in the photosynthetic membrane. *FEBS Lett.* 169, 1-6 (1984).
- Barros, J. et al. Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nat. Plants* 2, 16050 (2016).
- Bonner, C. a et al. Cohesion group approach for evolutionary analysis of TyrA, a protein family with wide-ranging substrate specificities. *Microbiol. Mol. Biol. Rev.* *MMBR* 72, 13-53 (2008).
- Hudson, G., Wong, V. & Davidson, B. Chorismate mutase/prephenate dehydrogenase from *Escherichia coli* K12: purification, characterization, and identification of a reactive cysteine. *Biochemistry (Mosc.)* 23, 6240-6249 (1984).
- Fischer, R. S. & Jensen, R. A. Prephenate dehydrogenase (monofunctional). *Methods Enzymol.* 142, 503-507 (1987).
- Rippert, P. & Matringe, M. Purification and kinetic analysis of the two recombinant aroenate dehydrogenase isoforms of *Arabidopsis thaliana*. *Eur. J. Biochem.* 269, 4753-4761 (2002).
- Dal Cin, V. et al. Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *Plant Cell* 23, 2738-53 (2011).
- Dornfeld, C. et al. Phylobiochemical characterization of class-ib aspartate/prephenate aminotransferases reveals evolution of the plant aroenate phenylalanine pathway. *Plant Cell* 26, 3101-14 (2014).
- Maeda, H., Yoo, H. & Dudareva, N. Prephenate aminotransferase directs plant phenylalanine biosynthesis via aroenate. *Nat. Chem. Biol.* 7, 19-22 (2011).
- Gaines, C. G., Byng, G. S., Whitaker, R. J. & Jensen, R. A. L-Tyrosine regulation and biosynthesis via aroenate dehydrogenase in suspension-cultured cells of *Nicotiana glauca*. *Plant Physiol.* 156, 233-240 (1982).
- Connelly, J. A. & Conn, E. E. Tyrosine biosynthesis in *Sorghum bicolor*: isolation and regulatory properties of aroenate dehydrogenase. *Z. Naturforschung C J. Biosci.* 41, 69-78 (1986).
- Keller, B., Keller, E. & Lingens, F. Aroenate dehydrogenase from *Streptomyces phaeochromogenes*-purification and properties. *Biol. Chem. Hoppe. Seyler* 366, 1063-1066 (1985).

26. Mayer, E., Waldner-Sander, S., Keller, B., Keller, E. & Lingens, F. Purification of arogenate dehydrogenase from *Phenyllobacterium immobile*. *FEBS Lett.* 179, 208-212 (1985).
27. Rubin, J. L. & Jensen, R. A. Enzymology of L-tyrosine biosynthesis in mung bean (*Vigna radiata* [L.] Wilczek). *Plant Physiol.* 64, 727-734 (1979).
28. Gamborg, O. L. & Keeley, F. W. Aromatic Metabolism in Plants I. A study of the Prephenate Dehydrogenase from Bean Plants. *Biochim. Biophys. Acta* 115, 65-72 (1966).
29. Legrand, P. et al. Biochemical characterization and crystal structure of Synechocystis arogenate dehydrogenase provide insights into catalytic reaction. *Structure* 14, 767-776 (2006).
30. Song, J., Bonner, C. a, Wolinsky, M. & Jensen, R. a. The TyrA family of aromatic-pathway dehydrogenases in phylogenetic context. *BMC Biol.* 3, 13 (2005).
31. Christendat, D., Saridakis, V. & Turnbull, J. Use of site-directed mutagenesis to identify residues specific for each reaction catalyzed by chorismate mutase-prephenate dehydrogenase from *Escherichia coli*. *Biochemistry (Mosc.)* 37, 1573-1580 (1998).
32. Christendat, D. & Turnbull, J. L. Identifying groups involved in the binding of prephenate to prephenate dehydrogenase from *Escherichia coli*. *Biochemistry (Mosc.)* 38, 4782-4793 (1999).
33. Sun, W., Singh, S., Zhang, R., Turnbull, J. L. & Christendat, D. Crystal structure of prephenate dehydrogenase from *Aquifex aeolicus*: insights into the catalytic mechanism. *J. Biol. Chem.* 281, 12919-12928 (2006).
34. Sun, W. et al. The crystal structure of *Aquifex aeolicus* prephenate dehydrogenase reveals the mode of tyrosine inhibition. *J. Biol. Chem.* 284, 13223-13232 (2009).
35. Lütke-eversloh, T. & Stephanopoulos, G. Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of *Escherichia coli*: generation and characterization of tyrosine-insensitive mutants. *Appl. Environ. Microbiol.* 71, 7224-7228 (2005).
36. Wierenga, R. K., De Maeyer, M. C. H. & Hol, W. G. J. Interaction of pyrophosphate moieties with alpha-helices in dinucleotide-binding proteins. *Biochemistry (Mosc.)* 24, 1346-1357 (1985).
37. Rippert, P., Puyaubert, J., Grisolle, D., Derrier, L. & Matringe, M. Tyrosine and phenylalanine are synthesized within the plastids in Arabidopsis. *Plant Physiol.* 149, 1251-60 (2009).
38. Jung, E., Zamir, L. O. & Jensen, R. A. Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. *Proc. Natl. Acad. Sci. U.S.A* 83, 7231-7235 (1986).
39. Wang, M., Toda, K. & Maeda, H. A. Biochemical properties and subcellular localization of tyrosine aminotransferases in *Arabidopsis thaliana*. *Phytochemistry* 132, 16-25 (2016).
40. Westfall, C. S., Xu, A. & Jez, J. M. Structural evolution of differential amino acid effector regulation in plant chorismate mutases. *J. Biol. Chem.* 289, 28619-28 (2014).
41. Cardoso, D. et al. Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. *Am. J. Bot.* 99, 1991-2013 (2012).
42. Wojciechowski, M. F., Lavin, M. & Sanderson, M. J. A phylogeny of legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. *Am. J. Bot.* 91, 1846-62 (2004).

43. Reyes-Prieto, A. & Moustafa, A. Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. *Sci. Rep.* 2, 955 (2012).
44. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691-9 (2001).
45. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30, 2725-2729 (2013).
46. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-1797 (2004).
47. Gorrec, F. The current approach to initial crystallization screening of proteins is under-sampled. *J. Appl. Crystallogr.* 46, 795-797 (2013).
48. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* 62, 859-866 (2006).
49. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* 64, 112-122 (2008).
50. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760-763 (1994).
51. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* 56, 965-972 (2000).
52. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* 374, 229-244 (2003).
53. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126-2132 (2004).
54. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213-221 (2010).
55. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658-674 (2007).
56. Maeda, H. et al. RNAi suppression of arogenate dehydratase1 reveals that phenylalanine is synthesized predominantly via the arogenate pathway in petunia petals. *Plant Cell* 22, 832-49 (2010).
57. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455-461 (2010).
58. Atkinson, H. J., Morris, J. H., Ferrin, T. E. & Babbitt, P. C. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4, e4345 (2009).
59. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498-2504 (2003).
60. Bonner, C. A., Fischer, R. S., Schmidt, R. R., Miller, P. W. & Jensen, R. A. Distinctive enzymes of aromatic amino acid biosynthesis that are highly conserved in land plants are also present in the chlorophyte alga *Chlorella sorokiniana*. *Plant Cell Physiol.* 36, 1013-1022 (1995).

Example 2—Conserved Molecular Mechanism of TyrA Dehydrogenase Substrate Specificity Underlying Alternative Tyrosine Biosynthetic Pathways in Plants and Microbes

In this Example, structure-guided phylogenetic analyses identified bacterial homologs, closely-related to plant TyrAs,

that also have an acidic 222 residue and ADH activity. A more distant archaeon TyrA that preferred PDH activity had a non-acidic Gln, whose substitution to Glu introduced ADH activity. Thus, the conserved molecular mechanism was involved in the evolution of arogenate-specific TyrAa in both plants and microbes.

This Example is based on data reported in Schenck et al., "Conserved Molecular Mechanism of TyrA Dehydrogenase Substrate Specificity Underlying Alternative Tyrosine Biosynthetic Pathways in Plants and Microbes," *Front Mol Biosci* 4:73 (2017), the contents of which (including all supplemental data, figures, and associated materials) is incorporated herein by reference.

#### Materials and Methods

##### Identification of Microbial TyrA Orthologs

BlastP® alignment tool searches were performed using the amino acid sequences of previously characterized TyrA homologs from plants (soybean PDH; GmPDH1 (Schenck et al., 2015) and Arabidopsis ADH; AtADH2; Rippert and Matringe, 2002) and microbes (*Synechocystis* sp. PCC6803 ADH (Legrand et al., 2006), and *E. coli* PDH (Hudson et al., 1984)) as the query in the NCBI database. This yielded only closely-related plant and microbial TyrA orthologs (e.g. algae and,  $\gamma$ -proteobacteria), which were then used as the query to perform additional BlastP® alignment tool searches. Every 5th BlastP® alignment tool hit was selected to provide sequences from various microbial lineages and limit bias in sample selection. Amino acid alignments were performed in PROMALS3D using the default parameters with structures of TyrA enzymes from plants and microbes with varying substrate specificities (*G. max* TyrAp; GmPDH1; PDB #5T8X, *H. influenzae* TyrAp81; HlPDH; 2PV7, and *Synechocystis* sp. PCC6803 TyrA<sub>2</sub>; SynADH; PDB #2F1K). Amino acid alignments from PROMALS3D were used to construct phylogenetic analyses using MEGA7. The analyses involved 130 amino acid sequences and all sites with less than 75% coverage were eliminated from the analysis. A neighbor-joining method was used to estimate evolutionary history using 1,000 bootstrap replicates (values shown at branches). The tree in FIG. 12 is a representative tree. Additional phylogenetic analyses were performed using the Maximum Likelihood method based on the Jones-Taylor-Thornton (JTT) matrix-based model, which gave overall similar results. All phylogenetic trees are drawn to scale, with branch lengths measured in the number of substitutions per site.

##### Recombinant Protein Expression and Purification and Site Directed Mutagenesis

Full length coding sequences from *Ochrobactrum intermedium* LMG 3301 (EEQ93947.1; OiTyrA), *Sediminispirochaeta smaragdinae* DSM 11293 (ADK80640.1; SsTyrA), and *Methanosaeta harundinacea* (KUK94425.1; MhTyrA) were optimized and inserted into pET28a vector using EcoRI and NdeI sites in frame with an N-terminal 6x-His tag.

For site directed mutagenesis of MhTyrA, plasmid template was diluted 100-fold, mixed with 0.04 U/ $\mu$ L Phusion DNA polymerase (Thermo), 0.2 mM dNTP's, 0.5  $\mu$ M forward

(5'-CATTCTGGCCGAAAGCCCG-GAACTGTATAGTAGC-3; SEQ ID NO: 167) and reverse (5'-GTTCCGGGCTTTCGGCCAGAATGCGGCC-CACAAAATC-3; SEQ ID NO: 168) mutagenesis primers, and 1x Phusion reaction buffer (Thermo), and then placed in a thermocycler for 98° C. for 30 s followed by 20 cycles of 10 s at 98° C., 20 s at 70° C., 4.5 min at 72° C. with a final extension at 72° C. for 10 min. The PCR products were purified with QIAquick Gel Extraction Kit (Qiagen), treated

with DpnI (Thermo) to digest methylated template DNA for 30 min at 37° C., and then transformed into *E. coli* XL1-Blue cells. Plasmids were sequenced to confirm that no errors were introduced during PCR and cloning.

For recombinant protein expression, *E. coli* Rosetta2 (DE3) cells (Novagen) transformed with the above plasmids were cultured as previously reported. For protein purification, 20 mL of the *E. coli* supernatant expressing the appropriate plasmid was applied to a 1 mL HisTrap FF column for purification of the His-tagged recombinant protein using an ÄKTA FPLC system (GE Healthcare). After loading the supernatant, the column was washed with 20 column volumes of 90% buffer A (0.5 M NaCl, 0.2 M sodium phosphate and 20 mM imidazole) and 10% buffer B (0.5 M NaCl, 0.2 M sodium phosphate and 0.5 M imidazole) followed by elution with 100% buffer B. Fractions containing purified recombinant enzymes were pooled and desalted by Sephadex G50 column (GE Healthcare) size-exclusion chromatography into lysis buffer. The purity of purified proteins were analyzed by SDS-PAGE using ImageJ software. All protein purification steps were performed at 4° C. unless stated otherwise.

##### ADH and PDH Assays

ADH and PDH assays were performed using purified recombinant enzymes for SsTyrA and MhTyrA Wt and Q227E mutant, while the *E. coli* cell lysate was used for OiTyrA as expression and purification of this enzyme was unsuccessful. Reactions contained 0.8 mM substrate (arogenate or prephenate) and 0.8 mM cofactor (NADP+ or NAD+) together with reaction buffer (25 mM HEPES pH 7.6, 50 mM KCl, 10% (v/v) ethylene glycol). For OiTyrA assays containing cell lysates, reactions were incubated for 45 minutes and analyzed using HPLC as previously reported (Schenck et al., 2015). For pure enzymes, reactions were monitored every 10-15 seconds for reduced cofactor at A340 nm using a microplate reader (Tecan Genios). Kinetic parameters of purified recombinant enzymes were determined from assays containing varying concentrations of arogenate (39.1  $\mu$ M-5 mM) or prephenate (39.1  $\mu$ M-5 mM) substrate and monitored 10-15 seconds for reduced cofactor at A340 nm using a microplate reader (Tecan Genios). Kinetic parameters were determined by fitting initial velocity data to the Michaelis-Menten equation using Origin software (OriginLab) from technical replicate assays (n=3). Arogenate substrate was prepared by enzymatic conversion of prephenate (Sigma-Aldrich). All enzyme assays were conducted at a reaction time and protein concentration that were in the linear range and proportional to reaction velocity.

##### Modeling Microbial TyrA Enzymes

Computation models were made using SWISS-MODEL with default parameters to predict the structures of divergent TyrA enzymes. Enzymes that are more closely-related to plants (e.g. SsTyrA and MhTyrA) were modeled using GmPDH1, though this resulted in a poor model for BdTyrA, which falls within the outgroup. BdTyrA was additionally modeled using *Synechocystis* sp. PCC6803 ADH. Homology models were visualized using PyMOL.

##### Results

##### Phylogenetic Relationship of Plant and Microbial TyrAs

Previous studies suggested that plant TyrAs are not derived from an eukaryotic ancestor or through cyanobacterial endosymbiosis because they are most similar to other microbes including some proteobacteria (Schenck et al., 2017; Bonner et al., 2008; Dornfeld et al., 2014; Reyes-Prieto and Moustafa, 2012); however, their precise origin was unclear. To resolve the phylogenetic relationship of

TyrA orthologs from divergent organisms including plants and microbes, here we performed structure-guided phylogenetic analyses using PROMALS3D to achieve alignment of TyrA orthologs with low sequence similarities (see methods) (Pei and Grishin, 2007). Three distinct clades were identified that contain: plant TyrAs together with those from algae, spirochaetes,  $\alpha$ - and  $\delta$ -proteobacteria (clade I, shaded blue in FIG. 12), TyrA orthologs from some archaea, fungi,  $\gamma$ -proteobacteria, and chloroflexi (clade II, shaded green), and TyrA orthologs from various microbes, which formed the outgroup and contains previously characterized microbial TyrA orthologs from *Synechocystis* sp. PCC 6803 and *Aquifex aeolicus* having very low sequence similarity (~30%) to plant TyrAs (clade III, FIG. 12). Interestingly, TyrAs from some spirochaetes lineages (some of which are known to cause harmful human diseases like Lyme disease) (Pritt et al., 2016) formed a subclade with plant and algae TyrAs within clade I using various phylogenetic methods (FIG. 12). These data suggest that Plantae TyrA may have been acquired through horizontal gene transfer (HGT) from an ancestor of one of these closely-related microbes. Microbial TyrA Orthologs Containing an Acidic 222 Residue 165 Prefer ADH Over PDH Activity

The amino acid sequence alignment of TyrAs showed that the Asp222 residue, which is conserved across plant TyrAa was also highly conserved in clade I (FIG. 12). On the other hand, most sequences in clade II, including some archaea TyrA, have a non-acidic Gln residue at the corresponding 222 position (FIG. 12), similar to legume TyrAp 169 enzymes (Schenck et al., 2017). Homology models of representative TyrA from clade I—*Arabidopsis thaliana* ADH (AtADH2, Plantae) (Rippert and Matringe, 2002) and *Sediminispirochaeta smaragdinae* DSM 11293 (SsTyrA, spirochaetes)—and clade II—*Methanosaeta harundinacea* (MhTyrA, archaea)—generated using GmPDH1 structure as the template indeed showed that their acidic and non-acidic residues, respectively, correspond to Asp222 in the active site of plant TyrA (FIG. 15). These data together suggest that TyrAs from clade I are likely arogenate-specific TyrAa enzymes, whereas more distantly-related microbial TyrAs from clade II are likely prephenate-specific TyrAp enzymes.

To experimentally test if TyrAs from clade I have ADH activity, representative TyrA orthologs from two distinct subclades of clade I, spirochaetes (SsTyrA) and  $\alpha$ -proteobacteria (*Ochrobactrum intermedium*; OiTyrA, FIG. 12), were expressed in *E. coli* as recombinant enzymes and biochemically characterized. SsTyrA and OiTyrA were chosen as they are located at key phylogenetic boundaries within clade I and contain residues required for cofactor binding and catalysis (FIG. 15). Purified SsTyrA recombinant enzyme showed ADH activity with a slight preference for NAD<sup>+</sup> over NADP184<sup>+</sup> cofactor; however, PDH activity was not detectable (FIG. 13A). Similarly, the *E. coli* cell lysate expressing OiTyrA had ADH but not PDH activity and strongly preferred NAD<sup>+</sup> over NADP186<sup>+</sup> cofactor (FIG. 13B), although the purification of OiTyrA was not successful due to low expression. These results demonstrate that microbial TyrA orthologs from clade I, which contain an acidic residue at the corresponding 222 position (FIG. 12), are arogenate specific TyrAa enzymes. An Archaeon TyrA Containing a Non-Acidic Residue Prefers PDH Over ADH Activity

To test if TyrA orthologs from clade II, which contain a non-acidic residue at the corresponding 222 position, are prephenate specific TyrAp 193 enzymes, a representative archaeon TyrA from *Methanosaeta harundinacea* (MhTyrA) was biochemically characterized. MhTyrA was chosen as no

TyrAs from its subclade of clade II have previously been characterized (FIG. 12). Also, MhTyrA is a monofunctional enzyme, while some archaea, fungi, and  $\gamma$ -proteobacteria orthologs in clade II are bifunctional and have a chorismate mutase enzyme domain (Hudson et al., 1984; Shlaifer et al., 2017). MhTyrA was expressed in *E. coli* and the recombinant enzyme was purified to homogeneity using affinity-chromatography (FIG. 16) and used for biochemical analyses. Unlike plant and microbial TyrAa 200 orthologs from clade I, MhTyrA showed strong PDH and very weak ADH activity (FIG. 13C). Interestingly, MhTyrA strongly preferred NADP<sup>+</sup> over NAD<sup>+</sup> cofactor (FIG. 13C), like plant TyrAs. These results suggest that TyrA orthologs from clade II that have a non-acidic residue at the corresponding 222 position are TyrAp 204 enzymes that strongly prefer prephenate over arogenate substrate. A single Q227E mutation introduces ADH activity in an archaeon TyrAp

To test if the non-acidic residue of MhTyrAp 206 at the corresponding 222 position (G1n227) is involved in substrate specificity, site-directed mutagenesis was performed on MhTyrAp 207 to replace Gln227 with glutamate and generate the MhTyrAp Q227E mutant. The 208 purified recombinant MhTyrAp Q227E enzyme (FIG. 16) showed decreased PDH activity with a substantial gain of ADH activity (FIG. 14, Table 6) without altering cofactor preference (FIG. 17).

TABLE 6

Kinetic analysis of MhTyrAp wild-type and Q227E mutant enzymes				
Enzyme	Substrate	$k_{cat}$ (s <sup>-1</sup> )	$K_m$ (mM)	$k_{cat}/K_m$ (mM <sup>-1</sup> s <sup>-1</sup> )
Wild-type	prephenate	2.44 ± 0.38	0.378 ± 0.02	6.44 ± 0.02
Wild-type	arogenate	N.D.	N.D.	N.D.
Q227E	prephenate	0.285 ± 0.17	2.669 ± 0.32	0.107 ± 0.01
Q227E	arogenate	0.704 ± 0.06	3.290 ± 0.22	0.213 ± 0.01

N.D. activity below detection limit

Kinetic analyses were conducted as described in FIG. 14 legend

Further kinetic analyses showed that wild-type MhTyrAp had a  $K_m$  of 378  $\mu$ M and turnover rate ( $k_{cat}$ ) of 2.4 s<sup>-1</sup> using prephenate substrate and NADP<sup>+</sup> 212 cofactor (FIG. 14, Table 6), which are comparable to previously characterized microbial TyrAp 213 enzymes. The very weak ADH activity of MhTyrAp 214 wild-type (FIG. 14, Table 6) precluded it from kinetic analysis using arogenate.

The Q227E mutant, on the other hand, exhibited almost 10-fold reduction in  $K_m$  for prephenate (2.4  $\mu$ M), while the catalytic efficiency ( $k_{cat}/K_m$ ) was reduced by 60-fold (0.1 vs. 6.4 mM<sup>-1</sup> s<sup>-1</sup>, FIG. 3 and Table 1). The Q227E mutant displayed substantial ADH activity compared with wild-type with a  $K_m$  of 3.3 mM, similar to that of Q227E for prephenate (2.7 mM, FIG. 14, Table 6) though still 10-fold higher than that of wild-type for prephenate (FIG. 14, Table 6) and other previously characterized TyrAa 221 enzymes. The Q227E mutant had roughly 2-fold higher catalytic efficiency with arogenate than with prephenate (0.2 vs. 0.1 mM<sup>-1</sup> s<sup>-1</sup>, FIG. 13). These results demonstrate that the single nonacidic to acidic mutation (Q227E) can shift the substrate preference of MhTyrAp 224 from prephenate to arogenate, suggesting that a single residue is responsible for substrate specificity of archaea TyrAp enzymes.

## Discussion

Previous studies suggest that microbes predominantly use a PDH-mediated pathway to synthesize Tyr, whereas plants

mainly use an ADH-mediated Tyr pathway. In this study, structure-guided phylogenetic analyses from diverse organisms identified ADH-like sequences in some bacteria, e.g. spirochaetes,  $\alpha$ - and  $\delta$ -proteobacteria, which form a monophyletic clade with plant TyrAs (FIG. 12). Biochemical characterization further demonstrated that TyrAs from spirochaetes and  $\alpha$ -proteobacteria indeed have ADH, but not PDH activity (FIGS. 13A, 13B). A native TyrA enzyme purified from the  $\alpha$ -proteobacteria *Phenylobacterium immobile*, which belongs to the same  $\alpha$ -proteobacteria genus found in clade I, was also previously shown to have ADH, but not PDH activity. Therefore, our study revealed that arogenate-specific TyrA<sub>a</sub> enzymes are more widely distributed in microbes than previously thought.

Previous evolutionary studies revealed that plant aromatic amino acid pathway enzymes are derived from a wide range of, and sometimes unexpected microbial origins. For example, plant shikimate kinase is most likely derived from cyanobacteria endosymbiosis whereas plant prephenate aminotransferase and arogenate dehydratase involved in Phe biosynthesis are sister to Chlorobi/Bacteroidetes orthologs. However, the evolutionary origin of plant TyrAs is currently unknown. TyrAs from some spirochaetes were more closely related to plant and algae TyrAs than other microbial TyrAs from clade I (FIG. 12) and, like Plantae TyrAa enzymes, had a conserved acidic residue at the corresponding 222 position. BlastP® alignment tool searches across different spirochaetes genomes showed that plant-like TyrAs are restricted to the order Spirocheatales, and absent in Leptospirales, Brevinematales, and Brachyspirales (FIG. 18). Thus, the current result suggests that the common ancestor of algae and plants acquired a TyrAa enzyme from a spirochaetes ancestor likely through a novel HGT event, rather than from an  $\alpha$ -proteobacteria through mitochondria symbiosis.

The archaeon MhTyrA from clade II preferred PDH over ADH activity (FIG. 13C) and had a non-acidic residue at the 222 position (FIG. 12). This is consistent with previously characterized clade II TyrA enzymes from  $\gamma$ -proteobacteria and fungi, which also preferred PDH over ADH activity though they belonged to distinct subclades (FIG. 12). As almost all TyrA sequences within clade II have a non-acidic residue (Gln or Asn) at the corresponding 222 position, except for Chloroflexi TyrAs (FIG. 12), they are likely prephenate-specific TyrA<sub>p</sub> enzymes. In plant TyrAs, an acidic residue at the corresponding 222 position confers ADH activity and when mutated to a non-acidic Gln, switches to PDH activity (Schenck et al., 2017). The reciprocal mutation (Gln to Glu) on MhTyrA<sub>p</sub> reduced PDH activity while introducing ADH activity (FIG. 14, Table 6). These data suggest that mutation of the non-acidic to an acidic residue at the corresponding 222 position played a key role in the evolution of arogenate-specific TyrA<sub>a</sub> 266 enzymes in microbes from clade I that gave rise to plant TyrAs.

The outgroup (clade III) appears to contain TyrA enzymes with both PDH and ADH activity. Homology models of a microbial TyrAs from the outgroup (e.g., *Bifidobacterium dentium* TyrA; BdTyrA) were compared to previously crystallized GmPDH1 and *Synechocystis* ADH to determine if the substrate specificity mechanism of TyrAs from clade I and II are also conserved in clade III TyrAs (FIG. 19). The global conformations of these divergent TyrA proteins from clade I and III are similar in structure, though there are some differences, such as additional  $\alpha$ -helices around the C-terminal dimerization domain (FIG. 19). All structures have conserved catalytic Ser101 and His124 that directly interact

with ring hydroxyl of arogenate and prephenate substrate (Schenck et al., 2017), suggesting that the key catalytic residues have been maintained across divergent TyrAs. However, the two loop regions surrounding and recognizing the substrate side chain, which contain the 222 residue and critical for substrate specificity (Schenck et al., 2017), are not well conserved in clade III as compared to clade I TyrAs (FIG. 19). This makes it difficult to confidently assign a corresponding residue in clade III TyrAs to the 222 position of clade I TyrAs (FIG. 12). Thus, clade III TyrAs likely use a different molecular mechanism(s) for their substrate specificity than plant and closely-related microbial TyrAs from clade I and II.

In conclusion, the current study revealed that arogenate-specific TyrAa enzymes evolved in some bacterial lineages, through the acquisition of an acidic residue at the 222 position, which later gave rise to the TyrAs of algae and land plants likely through a novel HGT event. More recently, the same residue was mutated back to a non-acidic residue uniquely in legume plants, which resulted in prephenate-specific TyrA<sub>p</sub> enzymes (Schenck et al., 2017). Thus, in the course of TyrA enzyme evolution, microbial TyrA<sub>p</sub> were converted into microbial TyrA<sub>a</sub> and then to legume-specific TyrA<sub>p</sub> by altering the same active site residue from a non-acidic to an acidic, and then back to a non-acidic residue. Previous studies proposed that the ubiquitous presence of the ADH-mediated Tyr pathway among photosynthetic organisms is to avoid futile cycling of tocopherol and plastoquinone biosynthesis from HPP. Identification of arogenate-specific TyrA among many non-photosynthetic microbes may require revisiting the biological significance of the ADH versus PDH-mediated Tyr biosynthetic pathways in diverse organisms. Given that arogenate and prephenate substrate specificity of TyrAs can be readily converted by a single residue (FIG. 14, Table 6) (Schenck et al., 2017), there must be significant selection pressure to maintain the acidic 222 residue and thus ADH activity in many organisms. The molecular mechanism and the key amino acid residue regulating the biochemical properties of diverse TyrAs also enables the optimization of Tyr biosynthesis via two alternative Tyr biosynthetic pathways in both plants and microbes for enhanced production of pharmaceutically important natural products derived from Tyr (e.g. morphine and vitamin E).

#### REFERENCES FOR EXAMPLE 2

- Bonner, C. A., Disz, T., Hwang, K., Song, J., Vonstein, V., Overbeek, R., et al. (2008). Cohesion group approach for evolutionary analysis of TyrA, a protein family with wide-ranging substrate specificities. *Microbiol. Mol. Biol. Rev. MMBR* 72, 13-53. doi:10.1128/MMBR.00026-07.
- Dornfeld, C., Weisberg, A. J., K C, R., Dudareva, N., Jelesko, J. G., and Maeda, H. A. (2014). Phylobiochemical characterization of class-Ib aspartate/prephenate aminotransferases reveals evolution of the plant arogenate phenylalanine pathway. *Plant Cell* 26, 3101-14. doi: 10.1105/tpc.114.127407.
- Hudson, G., Wong, V., and Davidson, B. (1984). Chorismate mutase/prephenate dehydrogenase from *Escherichia coli* K12: purification, characterization, and identification of a reactive cysteine. *Biochemistry* 23, 6240-6249.
- Legrand, P., Dumas, R., Seux, M., Rippert, P., Ravelli, R., Ferrer, J.-L., et al. (2006). Biochemical characterization and crystal structure of *Synechocystis* arogenate dehydrogenase provide insights into catalytic reaction. *Structure* 14, 767-776. doi:10.1016/j.str.2006.01.006.

- Reyes-Prieto, A., and Moustafa, A. (2012). Plastid-localized amino acid biosynthetic pathways of Plantae are predominantly composed of non-cyanobacterial enzymes. *Sci. Rep.* 2, 955-967. doi:10.1038/srep00955.
- Rippert, P., and Matringe, M. (2002). Purification and kinetic analysis of the two recombinant arogenate dehydrogenase isoforms of *Arabidopsis thaliana*. *Eur. J. Biochem.* 269, 4753-4761. doi:10.1046/j.1432-1033.2002.03172.x.

- Schenck, C. A., Chen, S., Siehl, D. L., and Maeda, H. A. (2015). Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nat. Chem. Biol.* 11, 52-57. doi:10.1038/nchembio.1693.
- Schenck, C. A., Holland, C. K., Schneider, M. R., Men, Y., Lee, S. G., Jez, J. M., et al. (2017). Molecular basis of the evolution of alternative tyrosine biosynthetic routes in plants. *Nat. Chem. Biol.* advance online publication. doi: 10.1038/nchembio.2414.

## SEQUENCE LISTING

The patent contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<https://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US11136559B2>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

We claim:

1. An engineered arogenate dehydrogenase polypeptide comprising at least 95% sequence identity to any one of the polypeptides of SEQ ID NOs: 10-55 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10, SEQ ID NOs:124-143 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10, SEQ ID NOs:149-151 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10, SEQ ID NOs:201-212 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10, or SEQ ID NOs:218-242 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10, wherein the engineered mutant arogenate dehydrogenase polypeptide has increased prephenate dehydrogenase activity as compared to an identical sequence with a D or E amino acid at a position corresponding to amino acid residue 220 of SEQ ID NO: 10.

2. The engineered polypeptide of claim 1, wherein the non-acidic amino acid residue is an asparagine (N) amino acid residue or a cysteine (C) amino acid residue.

3. The engineered polypeptide of claim 1, wherein the polypeptide has greater prephenate dehydrogenase activity than arogenate dehydrogenase activity.

4. The engineered polypeptide of claim 1, wherein the engineered polypeptide is selected from the group consisting of SEQ ID NO: 10 (MtnCADH D220C), SEQ ID NO: 11 (MtnCADH D220N), SEQ ID NO: 12 (AtADH2 D241N), SEQ ID NO: 13 (AtADH2 D241C), a polypeptide having at least 95% sequence identity to SEQ ID NO: 10 and comprising a cysteine (C) amino acid residue at position 220 of SEQ ID NO: 10, a polypeptide having at least 95% sequence

identity to SEQ ID NO: 11 and comprising an asparagine (N) amino acid residue at position 220 of SEQ ID NO: 11, a polypeptide having at least 95% sequence identity to SEQ ID NO: 12 and comprising an asparagine (N) amino acid residue at position 241 of SEQ ID NO: 12, and a polypeptide having at least 95% sequence identity to SEQ ID NO: 13 and comprising a cysteine (C) amino acid residue at position 241 of SEQ ID NO: 13.

5. A polynucleotide encoding the engineered polypeptide of claim 1.

6. A cell comprising the engineered polypeptide of claim 1.

7. The cell of claim 6, wherein the cell is a plant cell.

8. A method for increasing production of at least one product of the tyrosine or HPP pathways in a cell comprising introducing the engineered polypeptide of claim 1 into the cell.

9. The method of claim 8, wherein the cell is a plant cell.

10. The engineered polypeptide of claim 1, wherein the polypeptide has a sequence at least 95% identical to any one of SEQ ID NOs: 10-55 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10, SEQ ID NOs:124-143 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10, or SEQ ID NOs:149-151 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10.

11. The engineered polypeptide of claim 1, wherein the polypeptide has a sequence at least 95% identical to any one of SEQ ID NOs: 10-55 having a non-acidic amino acid at a position corresponding to residue 220 of SEQ ID NO:10.

\* \* \* \* \*