New Phytologist

Research

Are seven amino acid substitutions sufficient to explain the evolution of high L-DOPA 4,5-dioxygenase activity leading to betalain pigmentation? Revisiting the gain-of-function mutants of Bean et al. (2018)

M. Alejandra Guerrero-Rubio[1], Nathanael Walker-Hale[1], Rui Guo[1], Hester Sheehan[1], Alfonso Timoneda[1], Fernando Gandia-Herrero[2] and Samuel F. Brockington[1]

[1]Department of Plant Sciences, University of Cambridge, Tennis Court Road, CB2 3EA, Cambridge, UK; [2]Departamento de Bioquímica y Biología Molecular A, Unidad Docente de Biología, Facultad de Veterinaria, Regional Campus of International Excellence 'Campus Mare Nostrum', Universidad de Murcia, 30100 Murcia, Spain

Authors for correspondence:
Samuel F. Brockington
Email: sb771@cam.ac.uk

Fernando Gandía-Herrero
Email: fgandia@um.es

Received: 6 December 2022
Accepted: 27 January 2023

New Phytologist (2023) 239: 2265–2276
doi: 10.1111/nph.18981

Key words: betalains, caryophyllales, DODA, horizontal swapping, L-DOPA 4,5-dioxygenase, specialised metabolism.

Summary

This work revisits a publication by Bean et al. (2018) that reports seven amino acid substitutions are essential for the evolution of L-DOPA 4,5-dioxygenase (DODA) activity in Caryophyllales. In this study, we explore several concerns which led us to replicate the analyses of Bean et al. (2018).

Our comparative analyses, with structural modelling, implicate numerous residues additional to those identified by Bean et al. (2018), with many of these additional residues occurring around the active site of BvDODAα1. We therefore replicated the analyses of Bean et al. (2018) to re-observe the effect of their original seven residue substitutions in a BvDODAα2 background, that is the BvDODAα2-mut3 variant.

Multiple in vivo assays, in both Saccharomyces cerevisiae and Nicotiana benthamiana, did not result in visible DODA activity in BvDODAα2-mut3, with betalain production always 10-fold below BvDODAα1. In vitro assays also revealed substantial differences in both cataly-tic activity and pH optima between BvDODAα1, BvDODAα2 and BvDODAα2-mut3 proteins, explaining their differing performance in vivo.

In summary, we were unable to replicate the in vivo analyses of Bean et al. (2018), and our quantitative in vivo and in vitro analyses suggest a minimal effect of these seven residues in altering catalytic activity of BvDODAα2. We conclude that the evolutionary pathway to high DODA activity is substantially more complex than implied by Bean et al. (2018).

Introduction

Betalains are taxonomically restricted specialised pigments that in plants are unique to the flowering plant order Caryophyllales (Brockington et al., 2011; Timoneda et al., 2019). In the betalain biosynthetic pathway, a minimum of four enzymatic steps are required to proceed from tyrosine to stable betalain pigments, the yellow betaxanthins and violet betacyanins (Fig. 1). The key committed enzymatic step in betalain biosynthesis requires L-DOPA 4,5-dioxygenase activity (DODA), catalysing the conversion of L-3,4-dihydroxyphenylalanine (L-DOPA) to betalamic acid, the central chromophore of betalain pigments. In Caryophyllales, DODA activity is performed by an enzyme encoded by a LigB gene, which was initially characterised in Portulaca grandiflora (Christinet et al., 2004). Following characterisation of LigB in P. grandiflora, DODA activity has been either characterised or implicated in multiple LigB homologs across betalain-pigmented Caryophyllales (Sasaki et al., 2009; Zhao et al., 2011; Gandía-Herrero & García-Carmona, 2012; Harris et al., 2012; Hatlestad et al., 2012; Casique-Arroyo et al., 2014; Chung et al., 2015; Qingzhu et al., 2016; Imamura et al., 2018).

Phylogenetic analysis of the LigB gene lineage in Caryophyllales identified that a gene duplication occurred early in the evolution of the order, giving rise to two major clades of LigB genes, termed DODAα and DODAβ (Brockington et al., 2015). Consequently, all betalain-pigmented lineages of Caryophyllales contain at least two LigB genes, including one paralog from the DODAα lineage and one paralog from the DODAβ lineage. The function of DODAβ is unknown, but a number of lines of evidence suggest that, following this duplication, neofunctionalisation occurred within the DODAα lineage leading to the evolution of DODA activity (Brockington et al., 2015; Sheehan et al., 2020). However, evolutionary patterns within the DODAα lineage are complex (Sheehan et al., 2020). In addition to the DODAα/DODAβ duplication, there have been at least nine duplications in the DODAα lineage resulting in all betalain-pigmented lineages of

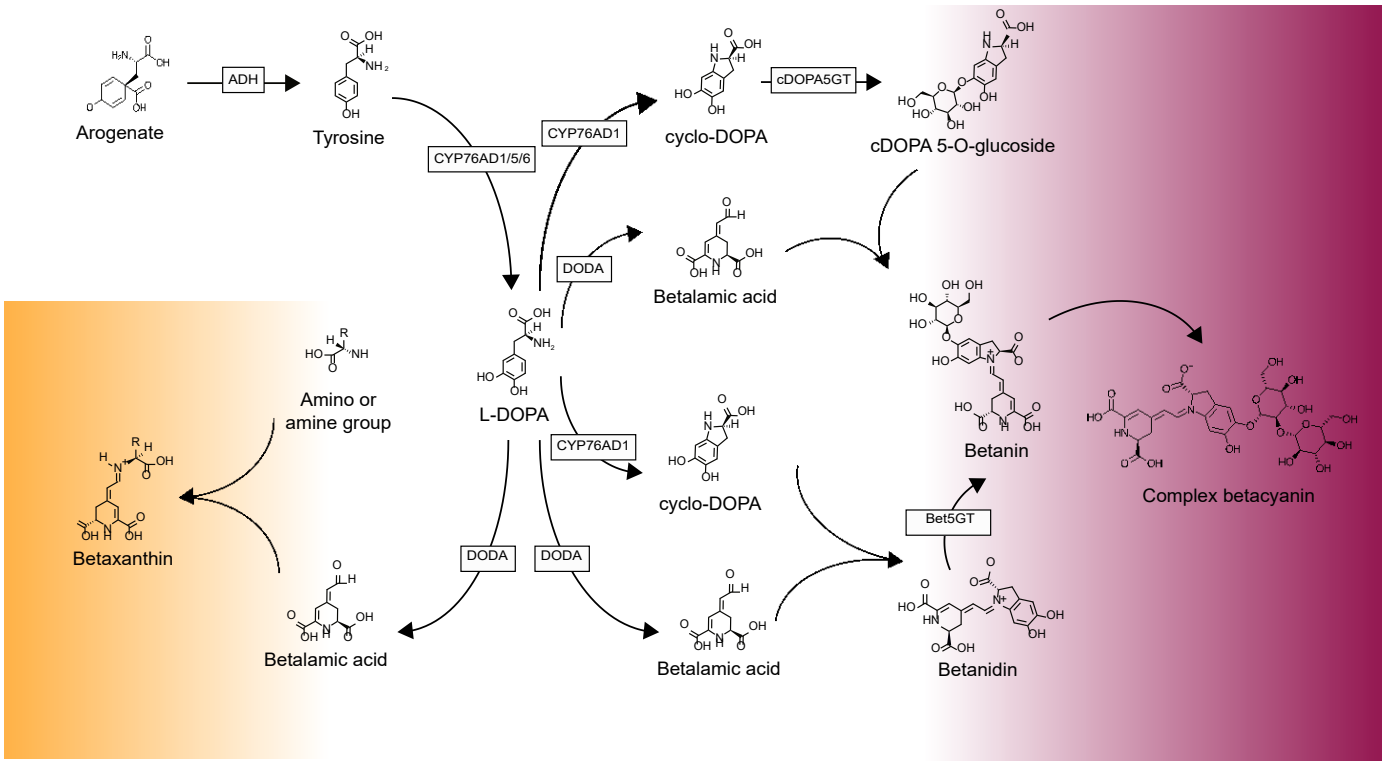

Fig. 1 Schematic showing the enzymatic and spontaneous reactions that form the specialised metabolites, betalains. The focus of this study is L-DOPA 4,5-dioxygenase (DODA), which catalyses the formation of betalamic acid, the core chromophore of betalain pigments, from L-3,4-dihydroxyphenylalanine (L-DOPA). Yellow shading corresponds to the formation of yellow pigments betaxanthins and pink shading corresponds to violet pigments betacyanins.

Caryophyllales containing at least three DODA genes – at least one homolog from the DODAβ lineage and at least two paralogs from the DODAα lineage, with Beta vulgaris containing five copies of DODAα (Sheehan et al., 2020). In B. vulgaris, two DODAα paralogs have been intensively studied (Sasaki et al., 2009; Gandía-Herrero & García-Carmona, 2012; Hatlestad et al., 2012; Chung et al., 2015; Bean et al., 2018), with one paralog, BvDODA1 (hereafter termed BvDODAα1), found to exhibit high levels of L-DOPA 4,5-dioxygenase activity and the other paralog, BvDODA2 (hereafter termed BvDODAα2), exhibiting marginal L-DOPA 4,5-dioxygenase activity.

Using this phenomenon of closely related paralogs BvDODAα1 and BvDODAα2 with different levels of L-DOPA 4,5-dioxygenase activity, Bean et al. (2018) used comparative analysis to identify the residues that are essential for L-DOPA 4,5-dioxygenase activity, using a horizontal swapping approach (Hochberg & Thornton, 2017). In this technique, a protein of interest with a particular biochemical function (in this case the L-DOPA 4,5-dioxygenase activity of BvDODAα1) is compared with a homologous protein that has identifiable similarity in sequence but a distinct function (in this case the absence of L-DOPA 4,5 cleavage in BvDODAα2). To identify the causal sequence differences for the functional variation, amino acid states in one protein are replaced with the corresponding states from the other, to identify the residue swaps that switch the function of one homolog (BvDODAα2) to the function of the homolog of interest (BvDODAα1; Fig. 2). Using this technique, Bean et al. report seven residues that, when altered in a BvDODAα2 background, were sufficient to allow BvDODAα2 to gain L-DOPA 4,5-dioxygenase activity on heterologous expression in Saccharomyces cerevisiae. Based on these data, Bean et al. concluded that 'seven amino acid mutations, are required for the B. vulgaris betalain-nonfunctional DODA paralog to evolve activity in L-DOPA ring cleavage'.

Theoretical concerns around the horizontal swapping approach (Hochberg & Thornton, 2017) led us to question the results of Bean et al. (2018), and the power of their experiments to explain the evolution of high L-DOPA dioxygenase activity. Furthermore, our own investigations of the system (Sheehan et al., 2020) led us to believe that many further residues were potentially implicated. We therefore attempted to replicate their analyses. We found that we could not justify their sole focus on the seven sites they selected, nor could we qualitatively or quantitatively replicate the results of their heterologous assays in S. cerevisiae, nor achieve comparable results in additional heterologous expression systems. To confirm this lack of replication and further explore these enzymes, we conducted in vitro enzymatic comparison of the two enzymes and the gain-of-function mutant, uncovering insights into enzyme differentiation in relation to high L-DOPA dioxygenase activity. Our results emphasise the difficulty of deriving evolutionary explanations from horizontal comparisons of extant enzymes and highlight further questions in understanding the evolution of DODA activity, in the context of betalain pigmentation.

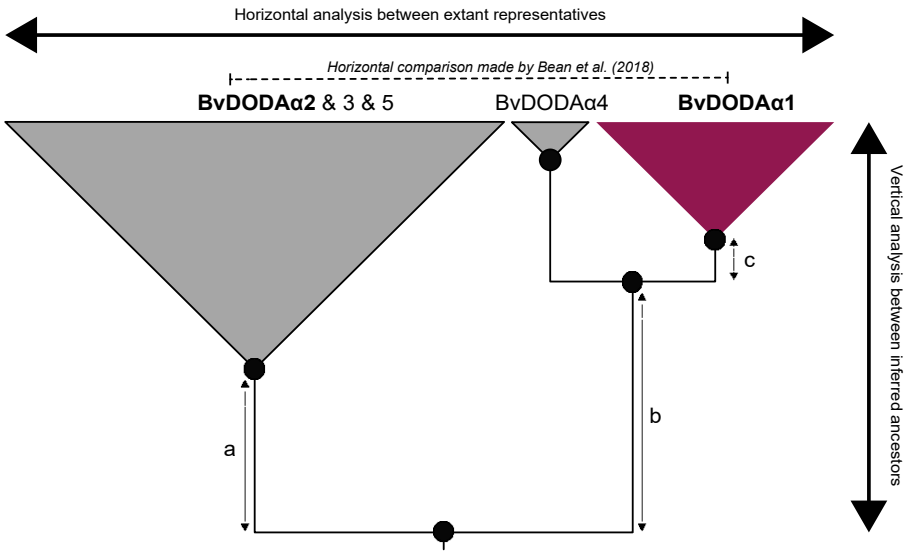

Fig. 2 Cladogram of the phylogenetic relationships between extant L-DOPA 4,5-dioxygenase (DODA) homologs in Beta vulgaris. Data derived from phylogenetic analysis in Sheehan et al. (2020), with clades collapsed into triangles proportional to number of sequences. Grey clades are not implicated in betalain pigmentation, and the purple clade is implicated in betalain pigmentation. Diagram depicts the difference between vertical approaches (sensu Hochberg & Thornton, 2017), vs the horizontal comparison made by Bean et al. (2018). Circles indicate inferred ancestors, and dotted lines (a, b, c) indicate the branch lengths that are relevant in understanding the evolution of function of BvDODAα1 relative to BvDODAα2. Note that BvDODAα4 shares a more immediate ancestor with BvDODAα1, than does BvDODAα2.

Materials and Methods

Reanalysis of comparative framework, diagnostic residues and structure

We collected the sequences listed in Bean et al. (2018) to replicate their comparative analysis. In doing so, we discovered two invalid accessions, meaning that we were unable to exactly reproduce their alignment; we replaced these with sequences from Brockington et al. (2015). We found that table S2 of Bean et al. (2018) did not include the two Caryophyllaceae nonbetalain sequences from Fig. 4(b) of Bean et al. (2018), so we included two candidates that matched the abbreviated binomial. Following Bean et al. (2018), we aligned the sequences using Clustal Omega v.1.2.4 (Sievers et al., 2011) with defaults and retained columns with at least 95% occupancy using pxclsq from Phyx v.1.1 (Brown et al., 2017; -p 0.95). We recovered a 229 amino acid alignment relative to the 225 amino acid alignment reported by (Bean et al., 2018). We inferred a maximum likelihood tree from the alignment using IQ-Tree v.2.1.2 (Nguyen et al., 2015) with the same model described by the authors (LG + Γ5) and 200 nonparametric bootstrap replicates, which was concordant with their tree. In doing so, we identified and corrected a labelling error in the initial Bean et al. (2018) table. Finally, we labelled sequences DODA1 or DODA2 according to Fig. 4(b) and table S2 of Bean et al. (2018). Corrected accessions and labels used to reproduce this analysis are available in the Supporting Information Table S1. We note that Bean et al. (2018) used DODA1 or DODA2 to refer both to specific sequences and to classes of sequences. Here, we refer to specific sequences by the nomenclature adopted in Sheehan et al. (2020; e.g. BvDODAα1) but refer to the two functional classes of sequences as DODA1 and DODA2 for clarity with Bean et al. (2018). To compare specific sequences, we calculated pairwise alignments with ClustalW as reported in the caption of Fig. S1 in Bean et al. (2018). Sequences, alignments and trees are available from https://github.com/NatJWalker-Hale/bean_response. To measure the divergence of sites according to DODA1 or DODA2 identity, we calculated a site-specific amino acid frequency vector each from sequences labelled DODA1 or DODA2 and calculated the base-2 Jensen–Shannon divergence (JSD) between the two vectors using the script calc_site_specific_divergence_aa.py available from https://github.com/NatJWalker-Hale/bean_response. This value is maximised at 1.0 when there is no overlap in amino acid states between the two classes, and 0 when all sequences have the same state across the two classes. We then ranked all residues according to JSD. To view the structural context of the residues selected by Bean et al. (2018), we predicted the protein structure of BvDODAα2 using the AlphaFold2 model as implemented in ColabFold v.1.4 (AlphaFold2_mmseqs2, https://github.com/sokrypton/ColabFold; Jumper et al., 2021; Mirdita et al., 2022). We coloured residues in the structure according to their JSD between DODA1 and DODA2 and viewed the structure in open-source Pymol v.2.5.0 (Schrodinger; https://www.schrodinger.com/products/pymol). We predicted the binding pocket of the structure using p2rank v.2.4, with the AlphaFold model (-c alphafold; Krivak & Hoksza, 2018).

Heterologous expression assay in Nicotiana benthamiana

Multigene binary vectors containing the genes of the betalain biosynthetic pathway (DODA, BvCYP76AD1 and MjcDOPA-5GT) were constructed using MoClo GoldenGate cloning following the protocol described (Engler et al., 2014; Timoneda et al., 2018). Transient expression using agroinfiltration of N. benthamiana was performed as described previously (Timoneda et al., 2018). The

controls in our experiment were as follows: positive, pBC-BvDODAα1; negative, uninfiltrated leaf. Samples were taken 4 d postinfiltration, betalains were extracted, and betanin was quantified using HPLC as described in Sheehan et al. (2020).

Heterologous expression assay in Saccharomyces cerevisiae by genomic integration

Saccharomyces cerevisiae BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was employed for the expression of BvCYP76AD6 and BvDODAs through their integration in the yeast genome by using Golden Gate Assembly and the yeast toolkit described in (Lee et al., 2015). For BvCYP76AD6, primers with cloning overhangs were used to amplify the coding sequence from Beta vulgaris L. ssp. vulgaris 'Bolivar' and the sequence has been deposited in GenBank (OQ362268). BvDODA coding sequences were synthesised as described with overhangs for cloning (Twist Bioscience, San Francisco, CA, USA). For all sequences, BsmBI and BsaI restriction sites were removed in the genes to facilitate plasmid construction and NotI restriction sites were removed to facilitate genomic integration. Cloning proceeded through two rounds: first, coding sequences were cloned into the part plasmid entry vector, pYTK001; second, coding sequences from the part plasmids were cloned into the integration vectors. BvCYP76AD6 was cloned into the URA3 integration vector, pYTK096, along with the ScTDH3 promoter (pYTK009) and the ScTHD1 terminator (pYTK056). BvDODA sequences were cloned into the LEU2 integration vector, pTMP137 (provided by J. E. Dueber, University of California, Berkeley, CA, USA), with the promoter ScCCW12 (pYTK010) and the terminator, ScADH1 (pYTK053). All plasmids were verified by Sanger sequencing and restriction digest. Plasmids were then linearised by digestion with NotI and transformed into yeast using the high-efficiency LiAc/SS carrier DNA/PEG yeast transformation protocol (Gietz & Schiestl, 2007). Via homologous recombination at the URA3 locus, the integration plasmid for BvCYP76AD6 was integrated into the genome of BY4741 to produce strain yHS023. All the integration plasmids for BvDODA were integrated into the genome of the yeast strain yHS023 via homologous recombination at the LEU2 locus. Cells were selected on complete supplement media lacking uracil or/and leucine. Genome integration was verified using primers specific to the URA3 or LEU2 integration junctions (Table S3; provided by J. E. Dueber, University of California, Berkeley). All the strains constructed in this work are listed in Table S2.

Heterologous expression assay in Saccharomyces cerevisiae by high plasmid copy strains

BvDODAα1, BvDODAα2 and BvDODAα2-mut3 sequences were also used as templates for their expression in yeast by an extrachromosomal, high-copy plasmid. Synthetically obtained fragments (Twist Bioscience) were codon-optimised and flanked by attB sites for their cloning by using Gateway cloning protocol where pDONR221 and pVV214 (URA3) were employed as donor vector and expression vector, respectively. Resulting destination vectors were verified by Sanger sequencing and employed for their expression in S. cerevisiae WAT11 (MATα (leu2Δ3,112 trp1Δ1 can1Δ100 ura3Δ1 ade2Δ1 his3Δ11,15)) using the 'Lazy bones' yeast transformation method (Burke et al., 2000). Positive colonies were selected on complete supplement media lacking uracil. Production of betalains was achieved by following Bean et al. (2018). Thus, transformed yeast were grown in complete supplement medium supplemented with galactose, 100 mg l^{1} leucine, 20 mg l^{1} histi-dine and 40 mg l^{1} adenine. Next day, cultures were pelleted by centrifugation to be resuspended at an $OD_{600\ nm}$ = 1.1 in fresh complete supplement medium containing the above-mentioned compounds and 10 mM L-DOPA and 2 mM ascorbic acid. All the strains constructed in this work are listed in Table S2.

Betalain quantification for Saccharomyces cerevisiae assays

Yeast strains grown in complete supplement media lacking uracil were also employed for quantification of their betalains content. Production of betalains by genomic integration was measured as follows: after 48 h of growing at 30°C, 14.49 g orbital shaking, 10 µl of saturated cultures were diluted into 490 µl of fresh medium supplemented with 1 mM tyrosine and 10 mM ascorbic acid and grown at 30°C, 120 g shaking in deep 96-well blocks. After 24 h, cells were centrifuged for 2 min at 2830 g and resuspended in phosphate-buffered saline (PBS) pH 7.4. After two rounds of washing, 100 µl of PBS-containing cells was used to quantify intracellular betaxanthin levels using a ClarioStar microplate reader (BMG Labtech, Ortenberg, Germany). Fluorescence of betaxanthins was detected by using excitation wavelength 470 nm and emission wavelength 510 nm. Values were normalised based on the negative control strain yHS023 and corrected by the cell density (OD_{600}). The same methodology was followed to produce betalains by extrachromosomal plasmids, but samples were supplemented with 10 mM L-DOPA and 2 mM ascorbic acid, as employed in Bean et al. (2018), and values were corrected by the cell density (OD_{600}).

Protein expression and purification

BvDODAα1, BvDODAα2 and BvDODAα2-mut3 sequences were used as templates to synthetically express them into the recombinant plasmid pGEX-4T-1. The new plasmids pGEX-BvDODAα1, pGEX-BvDODAα2 and pGEX-BvDODAα2-mut3 were purchased from Biomatik (Ontario, Canada), transformed into E. coli BL21 (Invitrogen) thermocompetent cells and plated onto LB agar plates containing ampicillin (Amp) 50 µg ml^{1}. E. coli cultures expressing pGEX-BvDODAα1, pGEX-BvDODAα2 and pGEX-BvDODAα2-mut3 were then employed for protein purification. Cells were grown at 37°C in 500 ml LB medium containing Amp 50 µg ml^{1} up to an OD_{600} = 0.8–1.2, when protein expression was induced by adding 0.2 mM IPTG. After 20 h at 20°C under orbital shaking, cells were harvested by centrifugation and resus-pended in PBS pH 7.4, supplemented with 0.5 mM iron (II) chlor-ide. Cell lysis was performed by sonication in a Cole-Parmer 4710 series ultrasonic homogeniser (Chicago, IL, USA). Recombinant DODA enzymes were then purified by Pierce™ Glutathione Agarose (Thermo Fisher, Waltham, MA, USA), according to the

manufacturer's instructions with the modification that PBS was employed throughout the protocol instead of the recommended buffer. After nontagged proteins were washed, the GST-tagged protein was eluted after addition of PBS supplemented with 0.1 mM reduced glutathione. Purified proteins were quantified using the Bradford assay (Bio-Rad; Bradford, 1976), and bovine serum albumin was used as standard to obtain a calibration curve. Samples were analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE), by application to 15% polyacrylamide gels and stained using a standard Coomassie Blue method.

Absorbance spectroscopy

Enzymatic ability to produce betalamic acid was determined using a continuous spectrophotometric method previously described by Gandía-Herrero & García-Carmona (2012). Briefly, a reaction media containing sodium phosphate buffer 50 mM, supplemented with 50 µM $FeCl_2$, L-DOPA 7.6 mM and sodium ascorbate 100 mM at a final volume of 300 µl, was employed to measure absorbance at λ = 414 nm. To standardise comparisons of the three enzymes, all enzymatic assays were performed with 500 ng ml^{1} of purified proteins. First, pH optima of the reaction were measured by the addition of 50 µl L-DOPA 7.6 mM to the reaction media (final concentration 1.3 mM) where different solutions of sodium phosphate buffer, from pH 5.5 to 8.5, were employed. Once the optimal pH was detected, different concentrations of L-DOPA 7.6 mM, ranging a final concentration from 0.125 to 3.8 mM, were added to the reaction media to measure the kinetic activity of the proteins at their optimal pH. Measurements were performed at 25°C in 96-well plates in a Synergy HT plate reader (Bio-Tek Instruments, Winooski, VT, USA). Betalamic acid solutions of known concentration were employed to calibrate the plate reader detector signal.

Trypsin digestion

Purified proteins were prepared in 100 µl of buffer NH_4HCO_3 50 mM, pH 8.0, with 0.02% ProteaseMAX™ Surfactant (Promega, Madison, WI, USA). Then, the samples were reduced with DTT 10 mM for 20 min at 56°C and alkylated with iodoacetamide 50 mM at room temperature in the dark for 20 min. One microgram of proteomics grade trypsin (Promega) was added, and the samples were incubated for 4 h at 37°C. Afterwards, samples were centrifuged at 15 000 g for 1 min to collect the condensate and the digestion was stopped by adding 0.5% TFA. Peptides were cleaned up with C18 Zip-Tips (Millipore) and evaporated using an Eppendorf vacuum concentrator model 5301.

Statistical analysis

Comparisons between wild-type sequences and mutants were conducted with Welch two-sample t-tests. Given that we aimed to test the hypothesis that BvDODAα2-mut3 increased activity over BvDODAα2, we tested against a one-sided alternative

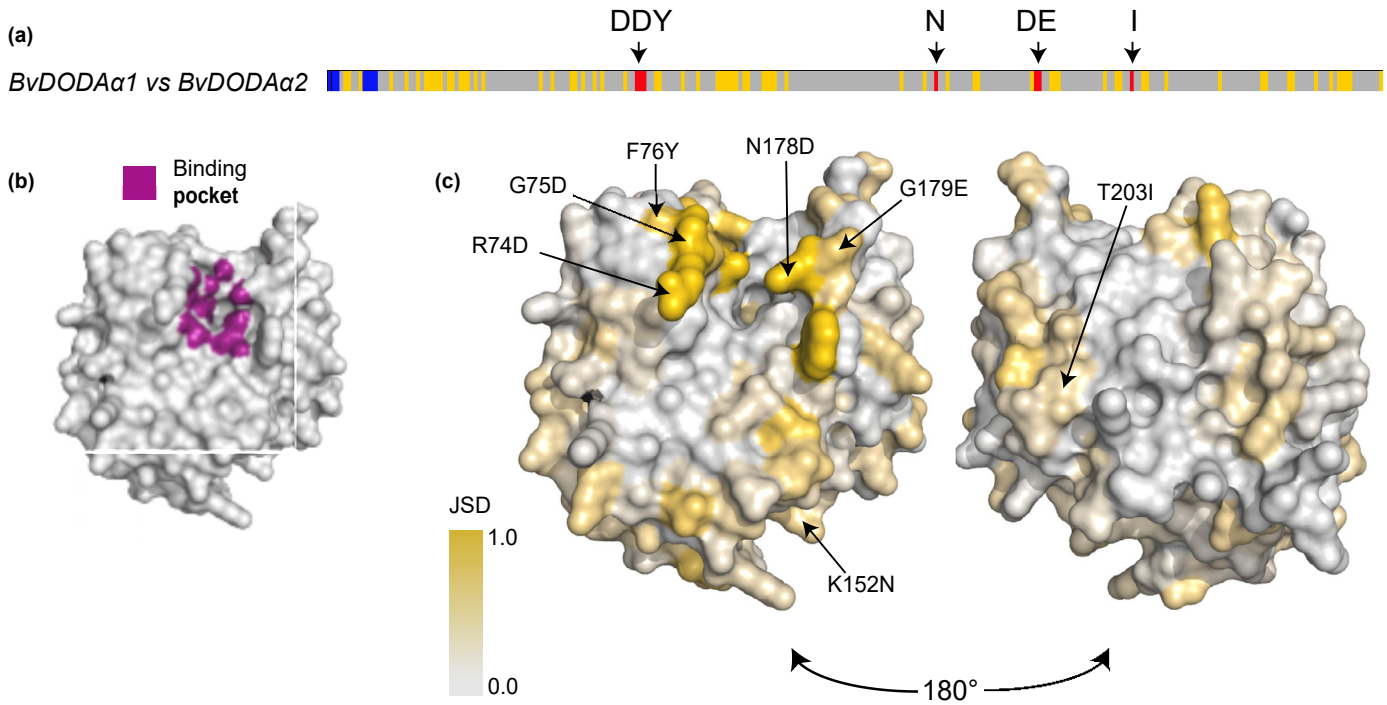

Fig. 3 Mutational differences between BvDODAα1 and BvDODAα2. (a) Sequence comparison between BvDODAα1 and BvDODAα2. Blue, indels relative to BvDODAα2; yellow, all substitutions; red, substitutions studied by Bean et al. (2018). (b) The predicted structure surface of BvDODAα2 from AlphaFold, highlighted with the predicted binding pocket from P2RANK. (c) The same structure coloured by the Jensen–Shannon divergence (JSD) between DODA1-and DODA2-like sequences studied by Bean et al. (2018). Left: facing binding pocket. Right: rotated 180°. 1.0 JSD indicates no overlap between the amino acid states of the two groups at that site, while 0.0 indicates identical distributions of amino acid states between the two groups. The seven mutations converting BvDODAα2 to BvDODAα2-mut3 are labelled.

hypothesis that the difference in means between BvDODAα2-mut3 and BvDODAα2 was >0, unless otherwise stated. Kinetic data analysis was performed using nonlinear regression fitted with the nls() function in R. Steady-state reactions fitted Michaelis–Menten equation and kinetic values from enzymes experiencing inhibition by excess of substrate were fitted with the corresponding equation described by Segel (1975) as follows:

$$v = \frac{V_{max}[S]}{K_m + [S] + \frac{[S]^2}{K_i}}$$

All analyses were conducted in R v.4.2.1 (R Development Core Team, 2016). Scripts and data for these analyses are available from https://github.com/NatJWalker-Hale/bean_response.

Results

Re-evaluating the seven residues in context

We re-examined the comparative analysis that ultimately led to focus on the seven amino acid residues reported by Bean et al. (2018). They selected 'residues that appeared to diverge according to betalain-producing activity' for mutagenesis. But we note that BvDODAα1 and BvDODAα2 differ by 78 amino acid substitutions and two inferred indels (comprising a total of seven residues; Fig. 3a) We reproduced the comparative analysis used by Bean et al. (2018) to select their sites for mutagenesis. We ranked sites in the alignment by their divergence according to DODA1 or DODA2 identity, and the top 41 sites, ranked by Jensen–Shannon divergence (JSD), are shown in Table 1. The seven residues mutated by Bean et al. (2018) are highlighted and range from the top three most divergent positions to the 41st position. Notably, given the comparative dataset used by Bean et al. (2018), the divergences calculated here indicate that there are numerous residues that are equally or more diagnostic of their DODA1 and DODA2 categories, than many of the seven residues that Bean and colleagues chose for subsequent analysis, and which they collectively deemed to be sufficient by analysis of their BvDODA2-mut3 variant. Bean et al. (2018) further justified their selection by reference to the structural findings of Christinet et al. (2004), particularly aspartate-rich regions in BvDODAα2 positions 74–76 and 178–179. To further examine the structural context of the mutated residues, we predicted the structure of BvDODAα2 and its binding pocket and analysed residues according to their divergence between DODA1 and DODA2. Our analysis shows that many of the highly divergent residues are proximal to the binding pocket, including several not analysed by Bean et al. (2018). Some of the residues mutated by Bean et al. (2018) do lie in proximity of the binding pocket, but two of their mutated residues are relatively distant (Fig. 3b,c).

Heterologous expression assay in Nicotiana benthamiana

We first attempted to reproduce and quantify the gain in L-DOPA 4,5-dioxygenase activity in BvDODA2-mut3 variant as reported by Bean et al. (2018). In our first experiment, we employed previously published constructs in conjunction with transient transformation in Nicotiana benthamiana, measuring betanin production, as a proxy for L-DOPA 4,5-dioxygenase activity (Timoneda et al., 2019; Sheehan et al., 2020). We synthesised the BvDODA2-mut3 based on a BvDODAα2 (Sheehan et al., 2020), arising from a different variety of B. vulgaris than that used by Bean et al., with two amino acid differences. While both BvDODAα2 and BvDODAα2-mut3 significantly increased measurable betanin over background (Welch two-sample t-test, P = 0.0463 and P = 0.0073, Fig. 4a), we found no visible pigmentation in the

Table 1 Jensen–Shannon divergence (JSD) between site-specific amino acid frequency vectors for sequences labelled DODA1 or DODA2 in Bean et al. (2018).

Position (alignment)	Position (BvDODAα2)	JSD DODA1-like vs DODA2-like
9	17	1
66	74	1
67	75	1
165	178	1
211	226	1
11	19	0.87
18	26	0.86
22	30	0.84
68	76	0.8
130	143	0.8
90	102	0.77
97	110	0.71
186	200	0.69
166	179	0.66
131	144	0.53
100	113	0.49
206	221	0.49
222	237	0.49
26	34	0.49
20	28	0.48
91	103	0.47
55	63	0.47
164	177	0.46
170	183	0.46
139	152	0.46
169	182	0.46
6	14	0.44
61	69	0.44
59	67	0.43
182	196	0.43
201	215	0.43
149	162	0.42
34	42	0.42
86	98	0.42
171	185	0.39
32	40	0.39
13	21	0.37
19	27	0.36
185	199	0.36
50	58	0.36
189	203	0.35

Original alignment position, position in BvDODAα2 (BvDODA2 in Bean et al., 2018), and JSD are shown. The seven residues mutated by Bean et al. (2018) are highlighted in yellow. Only the top 41 sites ranked by JSD are shown.

BvDODAα2-mut3 infiltration, and half the betanin content in BvDODAα2-mut3 infiltrations relative to BvDODAα2, albeit no statistically significant difference (two-sided Welch two-sample t-test, P = 0.2476, Figs 4a, S1). BvDODAα1 infiltrations had a c. 113- and c. 234-fold increase in mean betanin content over BvDODAα2 and BvDODAα2-mut3, respectively (Welch two-sample t-test, P = 0.0017 and P = 0.0017, Fig. 4a). We reasoned that the results we found, vs those recovered by Bean et al. (2018), might be due to two reasons: (1) the different heterologous host systems, that is our use of N. benthamiana vs their use of S. cerevisiae; (2) the two amino acid differences between the Bean et al., BvDODA2-mut3 and our version BvDODAα2-mut3.

Heterologous expression assay in Saccharomyces cerevisiae by genomic integration

We then shifted our efforts to obtain betalain pigmentation in BvDODAα2-mut3 by expressing BvDODAα1, BvDODAα2 and BvDODAα2-mut3 in S. cerevisiae, to better replicate Bean et al. (2018). We therefore synthesised versions of BvDODAα2 (HQ656022.1) and BvDODA2-mut3, codon-optimised for S. cerevisiae, based on the published protein sequences in Bean et al. (2018). We transformed the sequences by employing genomic integration, into a S. cerevisiae strain containing the BvCYP76AD6 gene, a well-known cytochrome P450-type enzyme with tyrosine hydroxylase activity (Polturak et al., 2016; Sunnadeniya et al., 2016) which produces the 3-hydroxylation of tyrosine to produce L-DOPA (Fig. 1), and quantified betaxanthin fluorescence as a proxy for L-DOPA 4,5-dioxygenase activity. Previous studies have established this as a sensitive method to quantify betaxanthin levels (DeLoache et al., 2015; Savitskaya et al., 2019). Although BvDODAα2 and BvDODAα2-mut3 had mean fluorescence significantly above background (Welch two-sample t-test, P = 0.016 and P = 0.001, Fig. 4b), we found no visible evidence of betaxanthin pigmentation after tyrosine feeding in either the BvDODAα2 or BvDODAα2-mut3 cultures whereas the BvDODAα1 culture produced a bright yellow colour (Fig. 5a) and showed c. 27- and c. 22-fold increases in mean fluorescence over BvDODAα2 and BvDODAα2-mut3 cultures, respectively (Welch two-sample t-test, P < 0.0001 and P < 0.0001, Fig. 4b). This high fluorescence contrasted to BvDODA2-mut3, which showed a 1.2-fold increased mean fluorescence over BvDODAα2 that was not statistically significant (Welch two-sample t-test, P = 0.08556, Fig. 4b).

Heterologous expression assay in Saccharomyces cerevisiae by high plasmid copy

Given our inability to produce betalains through the genomic integration of the DODA genes in S. cerevisiae, we then sought to precisely replicate their heterologous expression as described in Bean et al. (2018) – the same plasmids, the same yeast strain and the same methodology were employed. We synthesised versions of BvDODAα1, BvDODAα2 (HQ656022.1) and BvDODAα2-mut3, codon-optimised for S. cerevisiae and flanked with attB site

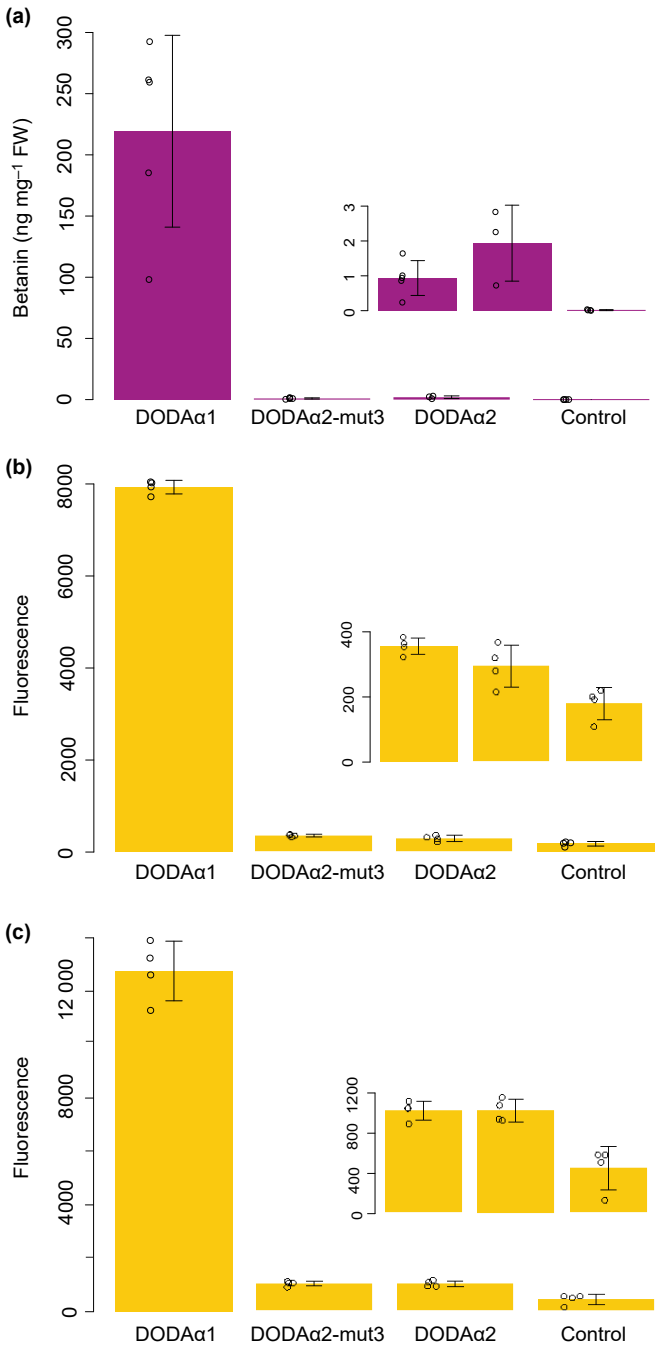

Fig. 4 Quantification of L-DOPA dioxygenase activity of BvDODAα1, BvDODAα2 and BvDODAα2-mut3 variants in heterologous backgrounds. (a) Quantification of betanin in Nicotiana benthamiana leaves after transient expression of multigene vectors that express BvDODAα1, BvDODAα2 and BvDODAα2-mut3 alongside BvCYP76AD1 and MjcDOPA-5GT. Bars show means from n = 5, 5, 3 and 4, 1 SD. (b) Fluorescence quantification of betaxanthins in Saccharomyces cerevisiae expressing BvDODAα1, BvDODAα2, or BvDODAα2-mut3, and BvCY-P76AD6 by genomic integration. Bars show means from n = 4, 1 SD. (c) Fluorescence quantification of betaxanthins in Saccharomyces cerevi-siae expressing BvDODAα1, BvDODAα2 and BvDODAα2-mut3 by high-copy, extrachromosomal plasmid. Bars show means from n = 4, 1 SD. Inserts on each graph magnifies the differences seen between BvDODAα2, BvDODAα2-mut3 and controls, which are otherwise difficult to see without rescaling.

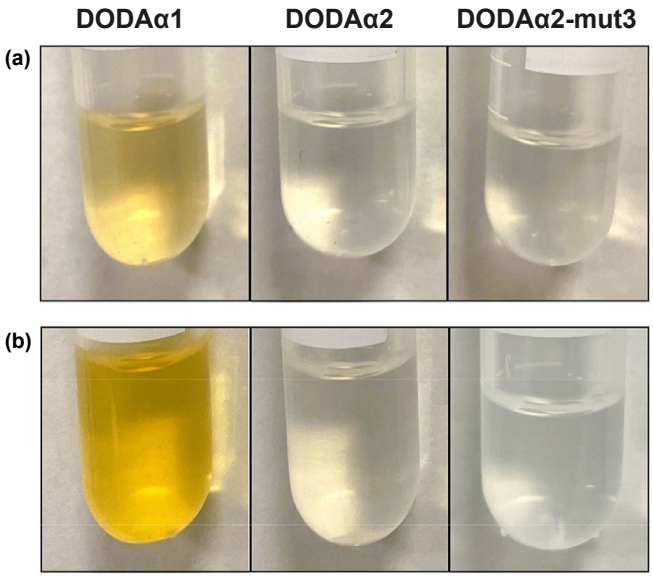

Fig. 5 Expression of BvDODAα1, BvDODAα2 and BvDODAα2-mut3 in Saccharomyces cerevisiae. (a) DODA expressed in S. cerevisiae BY4741 through genomic integration. Betalain accumulation was visible only in culture expressing BvDODAα1 when fed with 1 mM L-Tyr. No visible colouration was detected due to expression of BvDODAα2 and BvDODAα2-mut3. (b) Reproduction of the extrachromosomal expression of BvDODAs in S. cerevisiae WAT11 described in Bean et al. (2018). Only BvDODAα1 produced yellow colouration when fed with 10 mM L-DOPA.

for Gateway cloning, based on the published protein sequences in Bean et al. (2018). Our replication of the experiment however disagrees with the results previously indicated by Bean et al. (2018), since neither BvDODAα2 nor BvDODAα2-mut3 exhibited visible yellow coloration (Fig. 5b). Betalain content of each sample was then quantified by fluorescence and a marginal activity was detected in BvDODAα2. Both BvDODAα2 and BvDODAα2-mut3 had fluorescence significantly increased over the empty vector control (Welch two-sample t-test, $P = 0.0036$ and $P = 0.0039$, Fig. 4c). BvDODAα2-mut3 showed almost no difference in mean fluorescence to BvDODAα2 (1027.5 vs 1023.8, Welch two-sample t-test, $P = 0.4806$) and did not show bright yellow coloration to the naked eye (Figs 4c, 5b). BvDODAα1 meanwhile showed a c. 12.5-fold increase in mean fluorescence with respect to BvDODAα2 (Welch two-sample t-test, $P < 0.001$), and BvDODAα1 was the only sample with clearly visible bright yellow coloration (Figs 4c, 5b).

Protein purification of BvDODAα1, BvDODAα2 and BvDODAα2-mut3

We then sought to confirm our results through in vitro protein analysis. Difficulties in the purification process of plant DODA enzymes have previously been described by using His-tagged plasmids (Henarejos-Escudero et al., 2022). To solve this problem, a GST-tagged plasmid was employed in this study. Transformation of E. coli cells with pGEXT plasmids harbouring BvDODAα1, BvDODAα2 or BvDODAα2-mut3 allowed the purification of the protein through their heterologous expression, and the amount of protein recovered was measured employing Bradford method. To check homogeneity of these fractions, SDS-PAGE was employed. In BvDODAα2 and BvDODAα2-mut3, a single dominant band was detected (Fig. S2). The purification of BvDODAα1 yielded a similar dominant band but with a couple of additional much weaker bands. The dominant band from BvDODAα1 was extracted and characterised at molecular level through peptide mass fingerprint after trypsin digestion. Additionally, purified BvDODAa2 and BvDODAa2-mut3 were analysed by this same process. Main peptides from these samples are listed in Table S4. The peptides obtained from dominant BvDODAa1 band matched with the protein deposited under accession no. I3PFJ9 in Uniprot database, which corresponds to BvDODAα1, and peptides detected in BvDODAa2 and BvDODAa2-mut3 samples unequivocally confirmed their identity as in BvDODAa2 since both samples matched with the protein under accession no. A0A5B8XAF2.

In vitro enzymatic activity of BvDODAα1, BvDODAα2 and BvDODAα2-mut3

The in vitro activity of recombinant DODA enzymes was measured spectrophotometrically by the individual addition of each protein to a reaction medium with L-DOPA as a substrate. The addition of the enzyme yielded a yellow coloration with a kmax of 414 nm. No spectral change was detected in the absence of enzymes, showing that the change detected is due to the presence of the activity performed by the enzymes. The described activity agrees with the absorbance properties reported for those DODA enzymes previously employed in in vitro assays belonging to the fungus A. muscaria (Girod & Zryd, 1991), from the plants B. vulgaris (Gandía-Herrero & García-Carmona, 2012), M. jalapa (Sasaki et al., 2009) and Chenopodium quinoa (Henarejos-Escudero et al., 2022), the bacterium Gluconacetobacter diazotrophicus (Contreras-Llano et al., 2019) and the cyanobacterium Anabaena cylindrica (Guerrero-Rubio et al., 2020). Results showed that maximum activity of BvDODAα1 was detected at an optimal pH that differs to those detected for BvDODAα2 and BvDODAα2 mut3 (Figs 6a, S3). Maximum activity of BvDODAα1 was obtained at pH 6, while BvDODAα2 showed its highest activity at pH 8.5, agreeing with the optimal pH for BvDODAα2 previously described (Gandía-Herrero & García-Carmona, 2012). Optimum activity of BvDODAα2-mut3 with respect to pH was unchanged relative to BvDODAα2, at pH 8.5.

Different concentrations of L-DOPA were employed at the optimal pH of each protein to determine the kinetic parameters of these proteins (Fig. 6b,c). BvDODAα1 showed highest activity at low concentrations of L-DOPA that decreased as the concentration of L-DOPA increased (Fig. 6c). This phenomenon is named inhibition by excess of substrate (Segel, 1975) and has been reported for other betalain-forming DODA enzymes previously characterised (Contreras-Llano et al., 2019; Guerrero-Rubio et al., 2020). The equation for this kinetic model showed the parameters estimated (value 1 standard error) for BvDODAα1 as $K_m = 2.73$ 9.558 mM and $V_{max} = 47.658$ 158.372 μM min^{1} , and a strong substrate inhibition constant

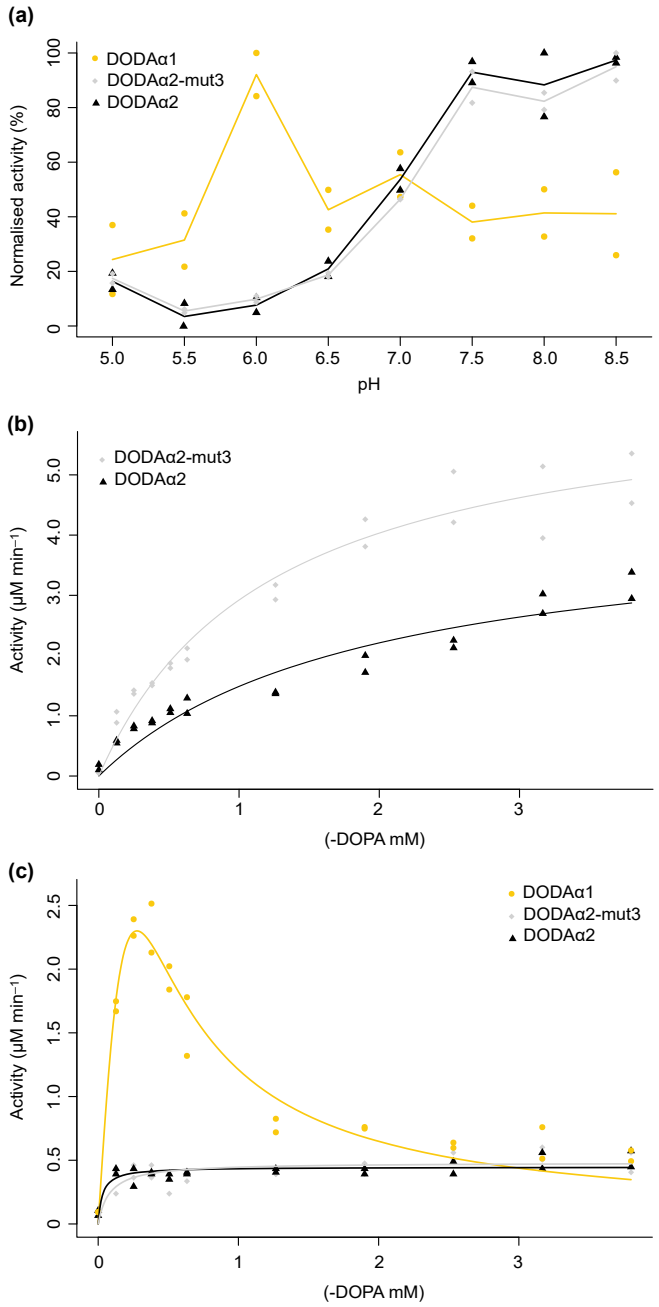

Fig. 6 In vitro characterisation of L-DOPA dioxygenase activity in BvDODAα1, BvDODAα2 and BvDODAα2-mut3. (a) Effect of pH on L-DOPA dioxygenase activity (n.b. activity has been normalised to % relative to the maximum of each enzyme). (b) Enzyme activity dependence on L-DOPA concentration measured in 50 mM sodium phosphate buffer, at pH 8.5 for BvDODAα2, and BvDODAα2-mut3. (c) Enzymatic activity dependence on L-DOPA concentration measured in 50 mM sodium phosphate buffer of BvDODAα1, BvDODAα2 and BvDODAα2-mut3 measured at pH 6.

(K_i) of 0.0281 0.0966 mM. Inhibition by excess of substrate was not detected in BvDODAα2 and BvDODAα2-mut3 activity and their results fitted a Michaelis–Menten curve (Fig. 6b). BvDODAα2-mut3 showed reduced K_m and a higher V_{max} (K_m = 1.228 0.197 mM and V_{max} = 6.511 0.418 μM min^{1}) relative to BvDODAα2 (K_m = 1.894 0.493 mM and V_{max} = 4.305 0.526 μM min^{1}). Due to the substrate inhibition of DODAα1, it is difficult to meaningfully compare kinetic parameters between BvDODAα1 vs BvDODAα2/BvDODAα2-mut3. But in any case, the characterisations show that kinetic behaviours are completely different. To emphasise the distinctive pH and substrate optima of BvDODAα1, the activity of BvDODAα2 and BvDODAα2-mut3 was also measured at pH 6 across a range of L-DOPA concentrations as depicted in Fig. 6(c).

Discussion

The phenomenon of betalain pigmentation is emerging as an important system in which to understand the origin and evolution of a novel metabolic pathway (Brockington et al., 2011). Previous studies have identified a host of mechanisms that underlie the evolution of the betalain biosynthesis pathway including seminal contributions by authors of Bean et al. (2018). These mechanisms include lineage-specific gene radiations (Brockington et al., 2015), duplication and neofunctionalisation (Brockington et al., 2015; Polturak et al., 2016; Sunnadeniya et al., 2016; Lopez-Nieves et al., 2018), co-option of biosynthetic enzymes (Vogt, 2002) and transcriptional regulators (Hatlestad et al., 2015), modification of primary metabolism (Lopez-Nieves et al., 2018; Timoneda et al., 2019) and indications of putative colinear gene clustering (Brockington et al., 2015; Sheehan et al., 2020). Given the role of duplication and neofunctionalisation, understanding the evolution of individual enzymes is an important piece of the evolutionary puzzle, with the potential to provide a rich mechanistic account of the evolutionary events by which proteins acquire new functions in betalain synthesis. Such lines of inquiry are even more interesting given the possibility of convergent specialisation to high L-DOPA 4,5-dioxygenase activity (Sheehan et al., 2020). However, the reconstruction of evolutionary paths to novel enzyme activity is complex (Hochberg & Thornton, 2017). In attempting to replicate the results of Bean et al. (2018), we have gained fresh insight into the properties of betalain biosynthetic enzymes and a clearer sense of the challenges in reconstructing the evolution of enzymatic activity leading to betalain biosynthesis.

The approach of horizontal swapping as applied by Bean et al. (2018) is intuitive, but suffers from well-documented flaws, and frequently fails to identify sequence differences that are necessary and sufficient for functional differences (Hochberg & Thornton, 2017). In short, there are two reasons for failure. First, horizontal comparisons are made against a background of all sequence differences between the extant homologs, which reflect all the changes that occurred along the lineages from the last common ancestor to the present-day proteins (Hochberg & Thornton, 2017; Fig. 2). Many or most of these changes will have nothing to do with the acquisition of the functional difference of interest (Bloom et al., 2007; Hochberg & Thornton, 2017). Second, horizontal swapping often produces nonfunctional proteins because of epistasis; that is, the phenotypic effects of a mutation are context dependent (Lunzer et al., 2010; Breen et al., 2012; Starr & Thornton, 2016). This may occur either because permissive residues required for the state to function are

absent in the recipient homolog or because restrictive residues that prevent it from functioning are present in the recipient homolog (Hochberg & Thornton, 2017). In either case, the consequence is that sequence differences that contribute to functional difference cannot be identified because their effect is masked by the presence or absence of other modifying residues (Hochberg & Thornton, 2017). In other words, a horizontal comparison, as attempted by Bean et al. (2018), should have a high theoretical chance of failure.

Bean et al. (2018) used a taxon-limited phylogenetic comparative approach to identify sites for mutagenesis, based on a comparison of residues diagnostic for what they term DODA1-like (betalain-functional) or DODA2-like identity (betalain nonfunctional). The authors use a single extant sequence, BvDODAα2, as an experimental background, and we note that BvDODAα2 contains 78 amino acid substitutions and two inferred indels (comprising a total of seven residues) compared with BvDODAα1 (Fig. 3a). We replicated their approach as closely as possible and detected many sites that were equally or more consistently divergent between the clades containing BvDODAα1 and BvDODAα2, than the majority of the seven sites they selected (Table 1). If consistent sequence divergence between respective clades and their conservation within these clades is the basis of testing residues, it seems unlikely that the seven sites identified alone should explain high L-DOPA 4,5-dioxygenase activity. Additional criteria, such as an understanding of the catalytic site and overall protein structure, can be informative in discriminating important sites. Indeed, Bean et al. (2018) refer to the regions implicated by Christinet et al. (2004) as important for L-DOPA binding activity in justifying their choices of sites to mutate. But many of the consistently divergent but experimentally untested sites seem to have as much potential to explain differences in activity as the seven residues reported by Bean et al. (2018), with several additional residues occurring around the predicted binding pocket and having high divergence. For example, alignment position 9/BvDODAα2 position 17 is predicted as a contact residue for the binding pocket in our analysis and is as consistently divergent as the most divergent sites selected by Bean et al. (2018; Table 1; Fig. 3b,c). All-in-all, it is unclear how these seven residues became the focus of Bean et al.'s study, to the exclusion of others.

Although the seven amino acid residues are referred to as 'key' or 'essential', the additional concept of sufficiency is implied by the images in Fig. 5 of Bean et al. (2018) that appear to show the same intensity of pigmentation in BvDODAα1 vs BvDODAα2-mut3. Here, we sought to replicate this increase of L-DOPA 4,5-dioxygenase activity in BvDODA2-mut3 vs BvDODAα2, but we were unable to replicate the visual gain in L-DOPA 4,5-dioxygenase activity in BvDODA2-mut3, nor to match the visible activity of BvDODAα1 (Fig. 5). Furthermore, the quantified L-DOPA 4,5-dioxygenase activity of BvDODAα1 was always an order of magnitude greater than BvDODA2-mut3 and BvDODAα2 in all heterologous experiments and was over 200-fold greater than BvDODA2-mut3 in Nicotiana benthamiana (Fig. 4). Given our inability to replicate the results of Bean et al. (2018), we argue that these seven residues are not sufficient to confer L-DOPA 4,5-dioxygenase activity in vivo to the levels seen with BvDODAα1. Whether all of these residues are essential is not fully evidenced by the qualitative horizontal swapping approach taken by Bean et al. (2018) and will depend on taking a quantitative vertical approach (Hochberg & Thornton, 2017), which analyses vertical residue changes and accounts for epistatic interactions.

Our in vitro data are consistent with our inability to replicate in vivo; that is, BvDODAα1 and BvDODAα2-mut3 bear little resemblance in terms of substrate concentration optima, substrate inhibition and pH optima. In our in vitro experiments, we see an increase in substrate affinity towards L-DOPA (K_m) in BvDODAα2-mut3 over BvDODAα2, and we also see that overall velocity of the reaction (V_{max}) is higher in BvDODAα2-mut3 than BvDODAα2 (Fig. 6b). However, overall, the seven residues do not modify the enzyme kinetics of BvDODAα2 to match the profile seen in BvDODAα1 (Fig. 6b,c). The substrate inhibition seen in BvDODAα1 is notably absent from BvDODAα2-mut3 and BvDODAα2, making it difficult to meaningfully compare K_m and V_{max} between BvDODAα1 and BvDODAα2/BvDODAα2-mut3 (Fig. 6b,c). Moreover, we need to be careful in drawing any major conclusions when comparing BvDODAα1 vs BvDODAα2/BvDODAα2-mut3, as BvDODAα2/BvDODAα2-mut3 exhibit only marginal activity and L-DOPA 4,5-dioxygenase activity is likely not their primary function. As BvDODAα2/BvDODAα2-mut3 lack substrate inhibition, they do eventually, at higher substrate concentrations, exhibit activity to the level seen in BvDODAα1; for example, the maximum uninhibited activity for BvDODAa1 (a maximum fitted value at 0.28 mM L-DOPA pH 6 in our model) is surpassed by BvDODAα2-mut3 (at 0.67 mM L-DOPA pH 8.5 based on our fitted model) or BvDODAα2 (at 2.17 mM L-DOPA pH 8.5 based on our fitted model). A further intriguing difference being that the pH optima of BvDODAα2 and BvDODAα2-mut3 remain the same at pH 8.5, while the pH optima for BvDODAα1 are much lower at c. pH 6 (Fig. 6a), where the activity of BvDODAα2/BvDODAα2-mut3 is negligible (Fig. 6c).

It is not necessarily clear how these in vitro substrate kinetics and pH optima translate in vivo. For example, it is difficult to predict how these differing pH optima of BvDODAα1 and BvDODAα2/BvDODAα2-mut3 play out in vivo, because while intracellular pH is strictly controlled, it is both dynamic and variable between different intracellular compartments. Betalains are most stable between pH 3–7 (Schwartz & von Elbe, 1983; Cai et al., 2001) and increasingly unstable above this range, and it seems likely that the betalain products could themselves degrade more rapidly at the higher pH optima of BvDODAα2/BvDODAα2-mut3. Traditionally betalain synthesis has been regarded as occurring in the cytosol where the pH is typically higher at c. pH 7.5, so the lower pH optima of pH 6 for DODAa1 is interesting and may suggest enzyme localisation in more acidic c. pH 6 intracellular compartments, such as multivesicular bodies, trans-Golgi networks and the vacuole (where betalains are stored; Grotewold, 2006; Shen et al., 2013). Likewise, there is little information on intracellular concentration of L-DOPA, so the differing optima for substrate concentration may also be informative in that respect; that is, the negligible activity of BvDODAα2/BvDODAα2-mut3 in vivo may indicate that

effective intracellular concentrations of L-DOPA are lower c. 0.3 mM. In effect, the contributions of a suboptimal pH and suboptimal substrate concentration are not possible to disentangle with any confidence, but it seems likely that both may contribute to low activity of BvDODAα2/BvDODAα2-mut3 in vivo.

In summary, we were unable to replicate the visible production of betaxanthin with BvDODAα2-mut3 reported by Bean et al. (2018). In heterologous in vivo assays, the activity of BvDODAα2-mut3 always remained at least 10-fold below that of wild-type BvDODAα1. Notably in Nicotiana benthamiana, the most physiologically relevant assays for these plant-derived DODA proteins, BvDODAα1 was on average over 200-fold more active than BvDODAα2-mut3. These in vivo discrepancies are supported by our in vitro analyses which indicate that the kinetic parameters of BvDODAα1 vs BvDODAα2/BvDODAα2-mut3 remain fundamentally different, which likely explains their differing in vivo performance in heterologous host platforms. We conclude that evolutionary path to BvDODAα1 activity remains substantially unsolved and is a more complex molecular and evolutionary challenge than implied by Bean et al. (2018).

Acknowledgements

We acknowledge support from the following funding bodies: SFB, BBSRC High Value Chemicals from Plants Network & NERC-NSF-DEB RG88096; HS, SNF P2BEP3_165359 & P300PA_174333; NWH, Woolf Fisher Cambridge Scholarship; RG, CSC no. (2018) 3101; MAGR, European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 101030560.

Competing interests

None declared.

Author contributions

MAGR, NWH, RG, HS and SFB planned and designed the work. MAGR, NWH, RG, HS and AT performed experiments and analysed the data. MAGR, NWH and SFB prepared the figures. SFB, MAGR, NWH, RG, HS, FGH and AT wrote the manuscript.

ORCID

Samuel F. Brockington https://orcid.org/0000-0003-1216-219X
Fernando Gandia-Herrero https://orcid.org/0000-0003-4389-3454
M. Alejandra Guerrero-Rubio https://orcid.org/0000-0002-3261-2058
Rui Guo https://orcid.org/0000-0002-5165-7905
Hester Sheehan https://orcid.org/0000-0002-2169-5206
Alfonso Timoneda https://orcid.org/0000-0002-7024-8947
Nathanael Walker-Hale https://orcid.org/0000-0003-1105-5069

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

References

Bean A, Sunnadeniya R, Akhavan N, Campbell A, Brown M, Lloyd A. 2018. Gain-of-function mutations in beet DODA2 identify key residues for betalain pigment evolution. New Phytologist 219: 287–296.

Bloom JD, Romero PA, Lu Z, Arnold FH. 2007. Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. Biology Direct 2: 17.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72: 248–254.

Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA. 2012. Epistasis as the primary factor in molecular evolution. Nature 490: 535–538.

Brockington SF, Walker RH, Glover BJ, Soltis PS, Soltis DE. 2011. Complex pigment evolution in the Caryophyllales. New Phytologist 190: 854–864.

Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, Wong GKS, Moore MJ, Smith SA. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. New Phytologist 207: 1170–1180.

Brown JW, Walker JF, Smith SA. 2017. Phyx: phylogenetic tools for unix. Bioinformatics 33: 1886–1888.

Burke D, Dawson D, Stearns T. 2000. Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual, 2000 edn. Plainview, NY, USA: Cold Spring Harbor Laboratory Press.

Cai Y, Sun M, Schliemann W, Corke H. 2001. Chemical stability and colorant properties of betaxanthin pigments from Celosia argentea. Journal of Agricultural and Food Chemistry 49: 4429–4435.

Casique-Arroyo G, Martínez-Gallardo N, González de la Vara L, Delano-Frier JP. 2014. Betacyanin biosynthetic genes and enzymes are differentially induced by (a)biotic stress in Amaranthus hypochondriacus. PLoS ONE 9: e99012.

Christinet L, Burdet FX, Zaiko M, Hinz U, Zrÿd J-P. 2004. Characterization and functional identification of a novel plant 4,5-extradiol dioxygenase involved in betalain pigment biosynthesis in Portulaca grandiflora. Plant Physiology 134: 265–274.

Chung H-H, Schwinn KE, Ngo HM, Lewis DH, Massey B, Calcott KE, Crowhurst R, Joyce DC, Gould KS, Davies KM et al. 2015. Characterisation of betalain biosynthesis in Parakeelya flowers identifies the key biosynthetic gene DOD as belonging to an expanded LigB gene family that is conserved in betalain-producing species. Frontiers in Plant Science 6: 499.

Contreras-Llano LE, Guerrero-Rubio MA, Lozada-Ramírez JD, García-Carmona F, Gandía-Herrero F. 2019. First betalain-producing bacteria break the exclusive presence of the pigments in the plant kingdom. mBio 10: e00345-19.

DeLoache WC, Russ ZN, Narcross L, Gonzales AM, Martin VJJ, Dueber JE. 2015. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nature Chemical Biology 11: 465–471.

Engler C, Youles M, Gruetzner R, Ehnert T-M, Werner S, Jones JDG, Patron NJ, Marillonnet S. 2014. A golden gate modular cloning toolbox for plants. ACS Synthetic Biology 3: 839–843.

Gandía-Herrero F, García-Carmona F. 2012. Characterization of recombinant Beta vulgaris 4,5-DOPA-extradiol-dioxygenase active in the biosynthesis of betalains. Planta 236: 91–100.

Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature Protocols 2: 31–34.

Girod P-A, Zryd J-P. 1991. Biogenesis of betalains: purification and partial characterization of dopa 4,5-dioxygenase from Amanita muscaria. Phytochemistry 30: 169–174.

Grotewold E. 2006. The genetics and biochemistry of floral pigments. Annual Review of Plant Biology 57: 761–780.

Guerrero-Rubio MA, García-Carmona F, Gandía-Herrero F. 2020. First description of betalains biosynthesis in an aquatic organism: characterization of

4,5-DOPA-extradiol-dioxygenase activity in the cyanobacteria Anabaena cylindrica. Microbial Biotechnology 13: 1948–1959.

Harris NN, Javellana J, Davies KM, Lewis DH, Jameson PE, Deroles SC, Calcott KE, Gould KS, Schwinn KE. 2012. Betalain production is possible in anthocyanin-producing plant species given the presence of DOPA-dioxygenase and L-DOPA. BMC Plant Biology 12: 34.

Hatlestad GJ, Akhavan NA, Sunnadeniya RM, Elam L, Cargile S, Hembd A, Gonzalez A, McGrath JM, Lloyd AM. 2015. The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. Nature Genetics 47: 92–96.

Hatlestad GJ, Sunnadeniya RM, Akhavan NA, Gonzalez A, Goldman IL, McGrath JM, Lloyd AM. 2012. The beet R locus encodes a new cytochrome P450 required for red betalain production. Nature Genetics 44: 816–820.

Henarejos-Escudero P, Martínez-Rodríguez P, Gómez-Pando LR, García-Carmona F, Gandía-Herrero F. 2022. Formation of carboxylated and decarboxylated betalains in ripening grains of Chenopodium quinoa by a dual dioxygenase. Journal of Experimental Botany 73: 4170–4183.

Hochberg GKA, Thornton JW. 2017. Reconstructing ancient proteins to understand the causes of structure and function. Annual Review of Biophysics 46: 247–269.

Imamura T, Takagi H, Miyazato A, Ohki S, Mizukoshi H, Mori M. 2018. Isolation and characterization of the betalain biosynthesis gene involved in hypocotyl pigmentation of the allotetraploid Chenopodium quinoa. Biochemical and Biophysical Research Communications 496: 280–286.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A et al. 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596: 583–589.

Krivák R, Hoksza D. 2018. p2rank: Machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. Journal of Cheminformatics 10: 39.

Lee ME, DeLoache WC, Cervantes B, Dueber JE. 2015. A highly characterized yeast toolkit for modular, multipart assembly. ACS Synthetic Biology 4: 975–986.

Lopez-Nieves S, Yang Y, Timoneda A, Wang M, Feng T, Smith SA, Brockington SF, Maeda HA. 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. New Phytologist 217: 896–908.

Lunzer M, Golding GB, Dean AM. 2010. Pervasive cryptic epistasis in molecular evolution. PLoS Genetics 6: e1001162.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. Nature Methods 19: 679–682.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32: 268–274.

Polturak G, Breitel D, Grossman N, Sarrion-Perdigones A, Weithorn E, Pliner M, Orzaez D, Granell A, Rogachev I, Aharoni A. 2016. Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants. New Phytologist 210: 269–283.

Qingzhu H, Chengjie C, Zhe C, Pengkun C, Yuewen M, Jingyu W, Jian Z, Guibing H, Jietang Z, Yonghua Q. 2016. Transcriptomic analysis reveals key genes related to betalain biosynthesis in pulp coloration of Hylocereus polyrhizus. Frontiers in Plant Science 6: 1179.

R Development Core Team. 2016. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL https://www.r-project.org/.

Sasaki N, Abe Y, Goda Y, Adachi T, Kasahara K, Ozeki Y. 2009. Detection of DOPA 4,5-dioxygenase (DOD) activity using recombinant protein prepared from Escherichia coli cells harboring cDNA encoding DOD from Mirabilis jalapa. Plant and Cell Physiology 50: 1012–1016.

Savitskaya J, Protzko RJ, Li F-Z, Arkin AP, Dueber JE. 2019. Iterative screening methodology enables isolation of strains with improved properties for a FACS-based screen and increased L-DOPA production. Scientific Reports 9: 5815.

Schwartz SJ, von Elbe JH. 1983. Identification of betanin degradation products. Zeitschrift Für Lebensmittel-Untersuchung Und Forschung 176: 448–453.

Segel IH. 1975. Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems. New York, NY, USA: John Wiley & Sons.

Sheehan H, Feng T, Walker-Hale N, Lopez-Nieves S, Pucker B, Guo R, Yim WC, Badgami R, Timoneda A, Zhao L et al. 2020. Evolution of L-DOPA 4,5-dioxygenase activity allows for recurrent specialisation to betalain pigmentation in Caryophyllales. New Phytologist 227: 914–929.

Shen J, Zeng Y, Zhuang X, Sun L, Yao X, Pimpl P, Jiang L. 2013. Organelle pH in the Arabidopsis endomembrane system. Molecular Plant 6: 1419–1437.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7: 539.

Starr TN, Thornton JW. 2016. Epistasis in protein evolution. Protein Science 25: 1204–1218.

Sunnadeniya R, Bean A, Brown M, Akhavan N, Hatlestad G, Gonzalez A, Symonds VV, Lloyd A. 2016. Tyrosine hydroxylation in betalain pigment biosynthesis is performed by cytochrome P450 enzymes in beets (Beta vulgaris). PLoS ONE 11: e0149417.

Timoneda A, Feng T, Sheehan H, Walker-Hale N, Pucker B, Lopez-Nieves S, Guo R, Brockington S. 2019. The evolution of betalain biosynthesis in Caryophyllales. New Phytologist 224: 71–85.

Timoneda A, Sheehan H, Feng T, Lopez-Nieves S, Maeda HA, Brockington S. 2018. Redirecting primary metabolism to boost production of tyrosine-derived specialised metabolites in planta. Scientific Reports 8: 17256.

Vogt T. 2002. Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from Dorotheanthus bellidiformis. Planta 214: 492–495.

Zhao S-Z, Sun H-Z, Gao Y, Sui N, Wang B-S. 2011. Growth regulator-induced betacyanin accumulation and dopa-4,5-dioxygenase (DODA) gene expression in euhalophyte Suaeda salsa calli. In Vitro Cellular & Developmental Biology – Plant 47: 391–398.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Representative HPLC chromatograms from transient expression of BvCYP76AD1, MjcDOPA-5GT and specified DODA variants in Nicotiana benthamiana leaves.

Fig. S2 SDS-PAGE electrophoretic analysis of (a) BvDODAα1, (b) BvDODAα2 and (c) BvDODAα2-mut3 from E. coli BL21.

Fig. S3 In vitro characterisation of L-DOPA dioxygenase activity in BvDODAα1, BvDODAα2 and BvDODAα2-mut3 showing the effect of pH on L-DOPA dioxygenase activity.

Table S1 Accessions used to reproduce the phylogenetic analysis from Bean et al. (2018).

Table S2 Information on constructed Saccharomyces cerevisiae strains.

Table S3 Information on primers employed in this work.

Table S4 Peptide mass fingerprint determined by MALDI-TOF analysis after trypsin digestion.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.