

# 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

## OPEN ACCESS

### Edited by:

Xiaoya Chen,  
Shanghai Institute of Plant Physiology  
and Ecology, China

### Reviewed by:

Michal Oren-Shamir,  
Volcani Center, Israel  
Guodong Wang,  
Institute of Genetics  
and Developmental Biology – Chinese  
Academy of Sciences, China

### \*Correspondence:

Kathy E. Schwinn,  
New Zealand Institute for Plant &  
Food Research Limited,  
Private Bag 11600,  
Palmerston North 4442, New Zealand  
kathy.schwinn@plantandfood.co.nz

### †Present address:

Hsiao-Hang Chung,  
Department of Horticulture, National  
Ilan University, 1, Sec. 1, Shen-Lung  
Road, I-Lan 260, Taiwan;  
Dion K. Harrison,  
Inno V8 Botanics, 12 Takara Court,  
Karana Downs, QLD 4306, Australia;  
Baxter Massey,  
3/21 Barlow Street, Scullin,  
ACT 2614, Australia;  
Hanh M. Ngo,  
11646 Sunlit Leaf Court, Houston,  
TX 77038, USA;  
Kate E. Calcott,  
Ministry for Primary Industries,  
P.O. Box 2526,  
Wellington 6140, New Zealand

### Specialty section:

This article was submitted to  
Plant Metabolism  
and Chemodiversity,  
a section of the journal  
Frontiers in Plant Science

Received: 23 April 2015

Accepted: 22 June 2015

Published: 07 July 2015

Hsiao-Hang Chung<sup>1†</sup>, Kathy E. Schwinn<sup>2\*</sup>, Hanh M. Ngo<sup>2†</sup>, David H. Lewis<sup>2</sup>,  
Baxter Massey<sup>1†</sup>, Kate E. Calcott<sup>2,3†</sup>, Ross Crowhurst<sup>4</sup>, Daryl C. Joyce<sup>1</sup>, Kevin S. Gould<sup>3</sup>,  
Kevin M. Davies<sup>2</sup> and Dion K. Harrison<sup>1†</sup>

<sup>1</sup> Centre for Native Floriculture, School of Agriculture and Food Sciences, The University of Queensland, Gatton, QLD, Australia, <sup>2</sup> New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand, <sup>3</sup> Victoria University of Wellington, Wellington, New Zealand, <sup>4</sup> New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand

Plant betalain pigments are intriguing because they are restricted to the Caryophyllales and are mutually exclusive with the more common anthocyanins. However, betalain biosynthesis is poorly understood compared to that of anthocyanins. In this study, betalain production and betalain-related genes were characterized in *Parakeelya mirabilis* (Montiaceae). RT-PCR and transcriptomics identified three sequences related to the key biosynthetic enzyme Dopa 4,5-dioxygenase (DOD). In addition to a LigB gene similar to that of non-Caryophyllales species (Class I genes), two other *P. mirabilis* LigB genes were found (*DOD* and *DOD-like*, termed Class II). *PmDOD* and *PmDOD-like* had 70% amino acid identity. Only *PmDOD* was implicated in betalain synthesis based on transient assays of enzyme activity and correlation of transcript abundance to spatio-temporal betalain accumulation. The role of *PmDOD-like* remains unknown. The striking pigment patterning of the flowers was due to distinct zones of red betacyanin and yellow betaxanthin production. The major betacyanin was the unglycosylated betanidin rather than the commonly found glycosides, an occurrence for which there are a few previous reports. The white petal zones lacked pigment but had DOD activity suggesting alternate regulation of the pathway in this tissue. *DOD* and *DOD-like* sequences were also identified in other betalain-producing species but not in examples of anthocyanin-producing Caryophyllales or non-Caryophyllales species. A Class I LigB sequence from the anthocyanin-producing Caryophyllaceae species *Dianthus superbus* and two *DOD-like* sequences from the Amaranthaceae species *Beta vulgaris* and *Ptilotus* spp. did not show DOD activity in the transient assay. The additional sequences suggests that DOD is part of a larger LigB gene family in betalain-producing Caryophyllales taxa, and the tandem genomic arrangement of two of the three *B. vulgaris* LigB genes suggests the involvement of duplication in the gene family evolution.

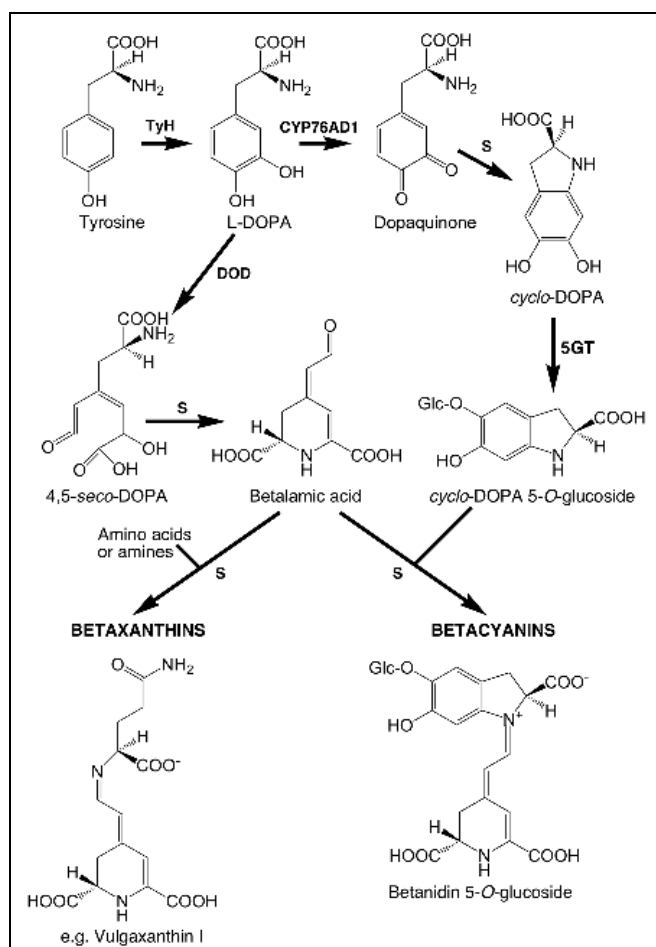
**Keywords:** dioxygenases, DODA, LigB, betalain, flower color, *Parakeelya*, pigmentation, Caryophyllales

## Introduction

Betalains, comprised of the red–violet betacyanins and yellow–orange betaxanthins, are plant pigments of special interest with regard to both their evolutionary origin and possible roles in stress-tolerance. They have a limited taxonomic distribution, having been found only in the flowering plant order Caryophyllales (Clement et al., 1994) and a few genera of the fungal phylum Basidiomycota. Betalains and the more common anthocyanin pigments have never been found together in the same plant (Mabry et al., 1963; Kimler et al., 1970), although betalains and other types of flavonoids do co-occur (Giannasi, 1988). For this reason, it is presumed that betalains, particularly the betacyanins, perform the same functions as anthocyanins, e.g., facilitating pollination and seed dispersal, and ameliorating abiotic stresses.

Although there has long been scientific interest in betalain pigments, there is much that remains to be elucidated. Notably, the evolution of this pathway and the mechanism behind the mutual exclusivity with anthocyanins has not been resolved (Sakuta, 2014). Intriguingly, there are a few families (Caryophyllaceae and Molluginaceae) within the core Caryophyllales that produce the more common, and extensively studied, anthocyanin pigments. However, as neither family are basal to the betalain families, and Molluginaceae actually resides within the betalain lineages, the betalain families are not monophyletic (Cuenoud et al., 2002; Brockington et al., 2009) and there is a complex evolutionary history of pigmentation in the order (Brockington et al., 2011). The biosynthetic and regulatory pathway for betalain production has also not been fully characterized. Until recently, only one of the enzymes responsible for formation of the colored compounds had been characterized at the molecular genetic level, dopa 4,5-dioxygenase (DOD). However, recent breakthroughs for *Beta vulgaris* (beet) have identified a gene for a second biosynthetic step (Hatlestad et al., 2012) and the first regulatory gene for the pathway, an R2R3MYB (Davies, 2015; Hatlestad et al., 2015).

In contrast to the phenylalanine-derived anthocyanins/flavonoids, betalains are derived from tyrosine. Betalamic acid (BA) is the chromophore of all plant betalains. In betalain biosynthesis (Figure 1), tyrosine is hydroxylated to form 3,4 dihydroxyphenylalanine (DOPA). A 4,5 aromatic ring cleavage of DOPA is then required to form BA, which proceeds through an unstable 4,5-*seco*-DOPA intermediate that spontaneously rearranges to form BA (Fischer and Dreiding, 1972; Schliemann et al., 1998). DOPA is also oxidized to *O*-DOPA-quinone, which spontaneously rearranges to form *cyclo*-DOPA, the key moiety that determines betacyanin formation. The conjugation reactions that form the pigments are thought to occur spontaneously (Strack et al., 2003; Harris et al., 2012) and involve *cyclo*-DOPA (or its glucoside) and BA for betacyanin synthesis, and BA and an amino acid or amine derivative other than *cyclo*-DOPA for betaxanthins. Thus, only three primary enzymatic reactions are thought to be involved in the pathway from tyrosine to the first colored compounds. Betalains are stored in the vacuole and betacyanins are normally stored as glucosides, with glucosylation possible at either the betanidin or *cyclo*-DOPA



**FIGURE 1 | Putative core betalain biosynthetic pathway in plants.**

S, spontaneous reaction; the primary enzymatic reactions are tyrosine hydroxylating activity (TyH), the enzyme for which has not been unequivocally determined; CYP76AD1, a cytochrome P450 enzyme that catalyzes oxidation of L-DOPA for the formation of *cyclo*-DOPA; and DOPA 4,5-dioxygenase (DOD) catalyzing the cleavage of L-DOPA needed for the formation of betalamic acid. Betacyanins normally accumulate as glucosides and may also be acylated. Regarding glucosylation at the 5-O-position, it appears that it is primarily done by *cyclo*-DOPA 5-O-glucosyltransferase (5GT) prior to the conjugation that forms the betacyanin.

steps of the biosynthetic pathway (Strack et al., 2003; Sakuta, 2014).

The biosynthetic enzymes for the latter two core steps have been characterized at the molecular level; DOD, which catalyzes DOPA ring cleavage, and CYP76AD(1), which forms *cyclo*-DOPA. CYP76AD has so far only been characterized from *B. vulgaris* (CYP76AD1, Hatlestad et al., 2012), *Mirabilis jalapa* (CYP76AD3, Suzuki et al., 2014) and *Amaranthus hypochondriacus* (CYP76AD1, Casique-Arroyo et al., 2014), but DOD has been studied across a few species. Gene sequences for DOD were first identified from the betalain producing fungus *Amanita muscaria* (Hinz et al., 1997) and subsequently in plants from moss rose (*Portulaca grandiflora*; Christinet et al., 2004). Plant DOD has homology to the LigB domain

of bacterial extradiol 4,5-dioxygenases, while the fungal DOD belongs to a distinct group of dioxygenases (Christinet et al., 2004). Despite their divergent sequences and distinct catalytic activity, both the plant and fungal DOD enzymes can function when transiently expressed in anthocyanin-producing species, as demonstrated by the associated accumulation of betalains when DOPA is supplied (Harris et al., 2012). LigB homologs have been identified in a wide range of plant orders (Christinet et al., 2004). However, those from non-betalain orders are readily distinguished by a distinctive sequence motif in the proposed catalytic domain (Christinet et al., 2004; Bahramnejad et al., 2010), and the *Arabidopsis* LigB (the only one for which a function has been assigned) is involved in the production of compounds not related to betalains, substituted  $\alpha$ -pyrones derived from phenylalanine (Weng et al., 2012). On this basis, we hereafter refer to them as Class I and Class II (containing DOD) type LigB homologs, respectively. The only report of a sequence for a Class I type LigB from a betalain-producing species is for *Amaranthus hypochondriacus* (*AhDODA-2*), which is referred to by the authors as being close to 'DODA-like' genes of anthocyanin-producing species (Casique-Arroyo et al., 2014).

While there is a great deal of information on the regulation of the expression of anthocyanin biosynthetic genes, there is little similar information for the betalain pathway. Although there are some studies showing DOD transcript levels correlating to betalain levels this seems to be variable between species (Wang et al., 2007; Casique-Arroyo et al., 2014), and studies on a wider range of species are required (Sakuta, 2014). In this study we examine betalain biosynthesis in a member of the Montiaceae family, *Parakeelya mirabilis*, by detailing betalain production, LigB Class II gene expression patterns and enzyme activities, and by generating a transcriptome database for the species. *Parakeelya* resides within the core Caryophyllales and its members (39–55 morphologically diverse species) are endemic to Australia (Carolin, 1993; Hershkovitz, 1998; Obbens, 2006). As a member of the Montiaceae, it is phylogenetically well separated from the model species *B. vulgaris*, which resides in the Amaranthaceae. Among the significant findings is the identification of two *P. mirabilis* LigB genes with distinct enzymatic activities and transcript abundance patterns, and we therefore also examine the occurrence and enzymatic activity of similar LigB genes from other members of the Caryophyllales. The results suggest previously unidentified LigB enzyme activities within the Caryophyllales, and also establish *Parakeelya* as a useful new model for studies to elucidate the betalain biosynthetic pathway.

## Materials and Methods

### Plant Material

An endemic Australian accession of *P. mirabilis*, supplied by Nindethana seed service (Albany, WA, Australia) was collected from Gascoyne, WA, Australia (25°02'586''S, 115°12'336''E). The genus was formed as a separation from the *Calandrinia* genus of the Portulacaceae family (Nyffeler and Eggli, 2010), and the species was formally known as *Calandrinia mirabilis*. The

lines were propagated by tissue culture as in Wickramasinghe et al. (2009). After 3 weeks, tissue culture plants were transplanted into potting medium of composted pine bark:sawdust:sand (2:4:1 v/v). The plants were grown in a greenhouse under ambient light conditions with drip irrigation at The University of Queensland, Gatton, QLD, Australia (27°33'272''S, 152°20'154''E). The average temperature and relative humidity (RH) were  $25 \pm 7^\circ\text{C}$  (range: 12–45°C) and  $70 \pm 29\%$  RH (range: 12–100%), respectively. Three genotypes (95-2, 146-2, and 149-1) were used as three biological replicates throughout the study. Floral developmental was divided into; stage 1 (closed bud approximately 0.5 cm long, no pigmentation), stage 2 (closed bud <1.0 cm long, pigmentation starting), stage 3 (closed bud <1.5 cm long, pale pigmentation), stage 4 (bud length <2.0 cm, strong pigmentation, petals extend from sepals), and stage 5 (bud length <3.0 cm, fully pigmented, open flowers). Stage 5 flowers were used for HPLC-LCMS analysis, mRNA extraction for transcriptome sequencing, and total RNA extraction for gene expression analysis of different plant tissues (including violet, white, red and yellow sectors of petals, stamens, stigmata, ovaries and leaves). Stage 4 flowers were used for gene cloning.

### Betalain Pigment Analyses

High-performance liquid chromatography as in Harris et al. (2012), with HPLC gradients as per Kugler et al. (2007), was used for betalain pigment identification and quantification in the different color sectors of *P. mirabilis* petals and the pigmented tissue from *PmDOD* transient expression experiments, with additional structural identification of pigments in the petal samples by LC-MS. Fresh tissue, or freeze dried tissue resuspended in 200  $\mu\text{L}$  deionised water (Svenson et al., 2008), was extracted in 80% (v/v) aqueous methanol containing 50 mM sodium ascorbate as in Schliemann et al. (1999). Peaks on the HPLC chromatograms were assessed against retention time and spectral maxima for known betalain pigments in a beetroot control sample, and further analyzed by LC-MS. The LC-MS system consisted of a Thermo Electron Corporation (San Jose, CA, USA) Finnigan Surveyor MS pump, Finnigan MicroAS auto-sampler, Finnigan Surveyor PDA detector and a Phenomenex ThermoSphere TS-130 column heater. The same mobile phase and elution gradients as for HPLC were applied. Two  $\mu\text{L}$  of extract was separated by reverse phase chromatography using a 4 mm  $\times$  2 mm 10  $\mu\text{m}$  Aqua guard cartridge and a 250 mm  $\times$  2.1 mm 4  $\mu\text{m}$  Phenomenex Synergi Hydro-RP C18 column at 25°C with a flow rate of 200  $\mu\text{L min}^{-1}$ . The eluent was scanned by PDA at 200–700 nm and analyzed by API-MS (Thermo-Finnigan LTQ, 2D linear ion-trap) with electrospray ionization (ESI, positive). Data were acquired for parent masses from  $m/z$  250 to 1500 amu with MS<sup>4</sup> fragmentation. Data were processed with the aid of Xcalibur®2.05 (Thermo-Electron Corporation).

A spectrophotometric method was used to measure total betalain content in *P. mirabilis* tissues. Fresh plant tissue (~100 mg) was homogenized to a fine powder in liquid nitrogen and extracted with 2 mL 80% (v/v) aqueous methanol

with 50 mM ascorbic acid (Schliemann et al., 1999). After centrifugation at  $15,000 \times g$  for 30 min at  $4^{\circ}\text{C}$ , the supernatant was quantified using a Pharmacia LKB-UltrospecIII UV/Vis spectrophotometer (Pharmacia LKB, Cambridge, UK). The aqueous pigment extracts were diluted with McIlvaine buffer (pH 6.0, citrate-phosphate) to obtain absorption values around 1.0 at absorption maxima. The betalain content (BC) was calculated after Kugler et al. (2004, 2007):  $[BC \text{ (mg kg}^{-1})] = (A - A_{650}) \times DF \times MW \times V_e \times 1000 / (\epsilon \times L \times W_d)$ ; where,  $A$  is absorbance value at the absorption maximum corrected by subtraction for non-specific absorbance at 650 nm,  $DF$  is the dilution factor,  $L$  is the cuvette path length (1 cm),  $V_e$  is total extract volume (mL), and  $W_d$  is dry weight of extracted material. Betacyanin were quantified using the MW ( $550 \text{ g mol}^{-1}$ ) and molar extinction coefficient ( $\epsilon = 60,000 \text{ L mol}^{-1} \text{ cm}^{-1}$  in water at  $\lambda_{\text{max}}$  538 nm) of betanin. Betaxanthins were quantified using the MW ( $339 \text{ g mol}^{-1}$ ) and molar extinction coefficient ( $\epsilon = 48,000 \text{ L mol}^{-1} \text{ cm}^{-1}$  in water at  $\lambda_{\text{max}}$  470 nm) of vulgaxanthin I. For conversion of betalain content from fresh weight (FW) to dry weight (DW), a different set of fresh tissue samples were collected and freeze dried for 1 week until completely dry. The weights of samples were measured before and after freeze-drying. The FW/DW ratio for three biological replicates was calculated and used for conversion.

### Isolation of cDNAs and Vector Construction

Based on homologous gene sequences obtained from publicly available databases, degenerate primers (Supplementary Table S1) were designed to amplify fragments of *LigB* homologs, *elongation factor 1 $\alpha$*  (*EF1 $\alpha$* ) and *actin* (*ACT*) from cDNA derived from *P. mirabilis* stage 4 petal RNA. Total RNA was extracted using an Isolate Plant RNA Mini Kit (Bioline, UK) and treated with DNase I (Promega, USA). First strand cDNA was made using a cDNA Synthesis Kit (Bioline, UK) with the dT<sub>18</sub> primers. The rest of the coding sequences were obtained by standard 3'- and 5'-rapid amplification of cDNA ends (RACE) polymerase chain reaction (PCR).

All vectors were based on pART7 (Gleave, 1992), which has a multiple restriction enzyme cloning site between the *Cauliflower Mosaic Virus* 35S (35S) promoter and terminator of the *Octopine Synthase* gene. For *PmDOD1*, *PmDOD-like*, and *PhybDOD-like* the cDNAs amplified started at the putative translation start sites and included part of the 3' untranslated regions (3' UTR). This amplified *PmDOD-like* had two amino acid changes [open reading frame (ORF) E192Q and T265S] compared to the original *PmDOD-like* sequence. *PhybDOD-like* was isolated from green leaf tissue of cv. Little Princess, a hybrid between *P. exaltatus* var. *semilanatus* B110A and *P. nobilis*. *BvDODA* was synthesized (GenScript USA Inc.) based on the ORF sequence from GenBank accession AJ583017. Synthesis incorporated restriction endonuclease sites flanking the putative translation start and stop codons to allow cloning into pART7. For *BvDODA1*, an ORF cDNA including stop codon was isolated from Swiss chard using primers specific for the GenBank sequence HQ656027. This cDNA was ligated into pGEMT-Easy and the *EcoRI* fragment then taken for vector construction. Plasmid preparations for transient assays used the AxyPrep<sup>TM</sup>

Plasmid Midiprep Kit (Axygen Biosciences). Primer sequences are given in Supplementary Table S1.

### Transient Gene Expression Assay

The assay used *Antirrhinum majus* line AG110 petal tissue and *Allium cepa* (onion) scales. *A. majus* line AG110 is a *nivea* mutant lacking chalcone synthase activity, and therefore has white flowers because no anthocyanins or other flavonoids are produced. The adaxial epidermis of each petal lobe sample was bombarded. Each experiment used three dorsal petal lobe sets per DNA/gold preparation, and each experiment was conducted at least twice. Each experiment compared a *DOD-like* construct with a *DOD* construct using the same biolistic conditions, with the latter also functioning as a positive control. Tissue surface sterilization, tissue culture and particle bombardment were as described in Shang et al. (2007), except for using a vacuum of  $-90 \text{ kPa}$ , a final construct DNA concentration of 1.0 or  $1.4 \mu\text{g DNA per } 1.0 \text{ mg gold particles}$ , and 4 or  $5 \mu\text{L}$  of gold suspension per shot. Each tissue sample was shot twice. The internal control to demonstrate transformation efficiency was pRT99GFP co-precipitated with the construct of interest at  $0.4 \mu\text{g DNA per } 1.0 \text{ mg gold particles}$ . After incubation overnight the bombarded tissues were vacuum infiltrated with sterilized water or either 10 or 1 mM L-DOPA (Sigma-Aldrich, USA), blotted (in some cases after a rinse with water to remove surface DOPA), and placed back on the medium for incubation overnight. Tissue was observed under white and blue (505–550 nm) light for the presence of pigments and fluorescence, respectively. Images were recorded using either an Olympus SZX fluorescent microscope (Olympus Corp., Tokyo, Japan) with a Leica DC500 digital camera (Leica Camera AG, Solms, Germany) or a Leica M205FA microscope with a Leica DFC550 digital camera. The assay using white onion scale was performed as for *A. majus*, except that surface sterilization was for 2 min and the solution did not contain Tween, the distance between the gene gun column and tissue was 12 cm, 1/2 MS medium was used with 1.2% (w/v) agar and 3% (w/v) sucrose, and tissue was incubated under  $40 \mu\text{mol m}^{-2} \text{ s}^{-1}$  light from Polylux 36 W fluorescent tubes (Polylux XL, Hungary) at room temperature. Tissue was observed under an Olympus BH2 light microscope (Olympus, Japan) and images recorded using a Canon IXUS400 digital camera (Tokyo, Japan).

### Quantitative Real-Time PCR (qPCR)

TaqMan primers and minor groove binder (MGB) probes labeled with 6-carboxy fluorescein (6FAM<sup>TM</sup>) are given in Supplementary Table S1. RNA extraction, DNase I treatment and cDNA synthesis were carried out as given already for isolation of cDNAs except that random hexamers were used as primers for reverse transcription. The total RNA used in each reverse transcription reaction was 500 ng. Quantitative (q)PCR was performed in a total reaction volume of  $25 \mu\text{L}$  (25 ng cDNA sample, 1x TaqMan<sup>®</sup> Universal PCR Master Mix (Life Technologies, Australia), 300 nM of each gene specific primer and 200 nM MGB probe) using an ABI7000 thermocycler (Applied Biosystems). Amplification was performed using the parameters of 2 min at  $50^{\circ}\text{C}$  and 10 min at  $95^{\circ}\text{C}$ , followed by 40

cycles of 15 s at 95°C and 30 s at 60°C. The cycle threshold (Ct) value for each reaction was calculated. Expression of the reference genes, *EF1α* and *ACT*, was consistent across development, with CT values changing by less than 2.0. Three biological replicates and three technical replicates for each biological replicate were performed. The natural logarithm transformed data were used for analysis. Expression of *PmDOD1* and *PmDOD2* was normalized by the BestKeeper Index of *EF1α* and *ACT* using BestKeeper® Software version 1 (Pfaffl et al., 2004).

### Maximum Likelihood Analysis

The ORF DNA from the LigB homologs were aligned using Muscle (Edgar, 2004) followed by Maximum Likelihood analysis using the PhyML program (Guindon and Gascuel, 2003) within the bioinformatics software Geneious version 6.1.6 (Biomatters, Auckland, New Zealand). The default parameters of substitution model set to LG, proportion of variable sites set to fixed 0, number of substitution rate categories set to 4, gamma distribution parameter set to estimated, optimize set to topology/length/rate were used. Topology search used either NNI (Default, fast) or BEST (Best of NNI and SPR search). Bootstrap analysis with 1000 replicates was used to estimate the confidence of each tree node.

### Transcript Sequencing and Assembly

Violet and yellow petal sectors were dissected, mRNA was isolated using the  $\mu$ MACS mRNA Isolation Kit (Miltenyi Biotec), and sent to the Australian Genome Research Facility (Brisbane, QLD, Australia) for 454 GS-FLX Titanium sequencing. *De novo* assembly of FASTA sequence was performed using MIRA3 (Chevreux et al., 1999) using default settings. Violet and yellow petal assemblies used 211,574 and 203,929 reads, respectively, with average read length of 321 and 323 bases, respectively. The yellow petal *de novo* assembly yielded 21,189 transcripts containing 11,943,110 bases with mean length of 563 bases, a maximum transcript length of 5,009 bases, and an assembly N50 of 603 bases. The violet petal *de novo* assembly yielded 22,662 transcripts containing 13,139,842 bases with mean length of 579 bases, a maximum transcript length of 8,136 bases and an assembly N50 of 620 bases. Data is available on the NCBI database linked to the Biosample accession numbers SAMN03769446 and SAMN03769447 for the violet and yellow samples, respectively.

## Results

### Betalain Biosynthesis in *Parakeelya*

Fully open flowers of *Parakeelya mirabilis* show four distinct color sectors of violet, white, red, and yellow (Figure 2). These were dissected and analyzed for betalain content (Figures 2 and 3), along with stamens, stigmata, ovaries, and leaves (Figure 2). It was not possible to avoid sampling a small amount of colored tissue when sampling the white sector tissue. The highest betalain content was detected in violet and red petal sectors followed by stigmata, yellow petal sectors and stamens. As was expected from the tissue dissection, a small amount of pigment was also detected in white petal sectors. Total betalain content increased markedly during petal development, with a major increase occurring

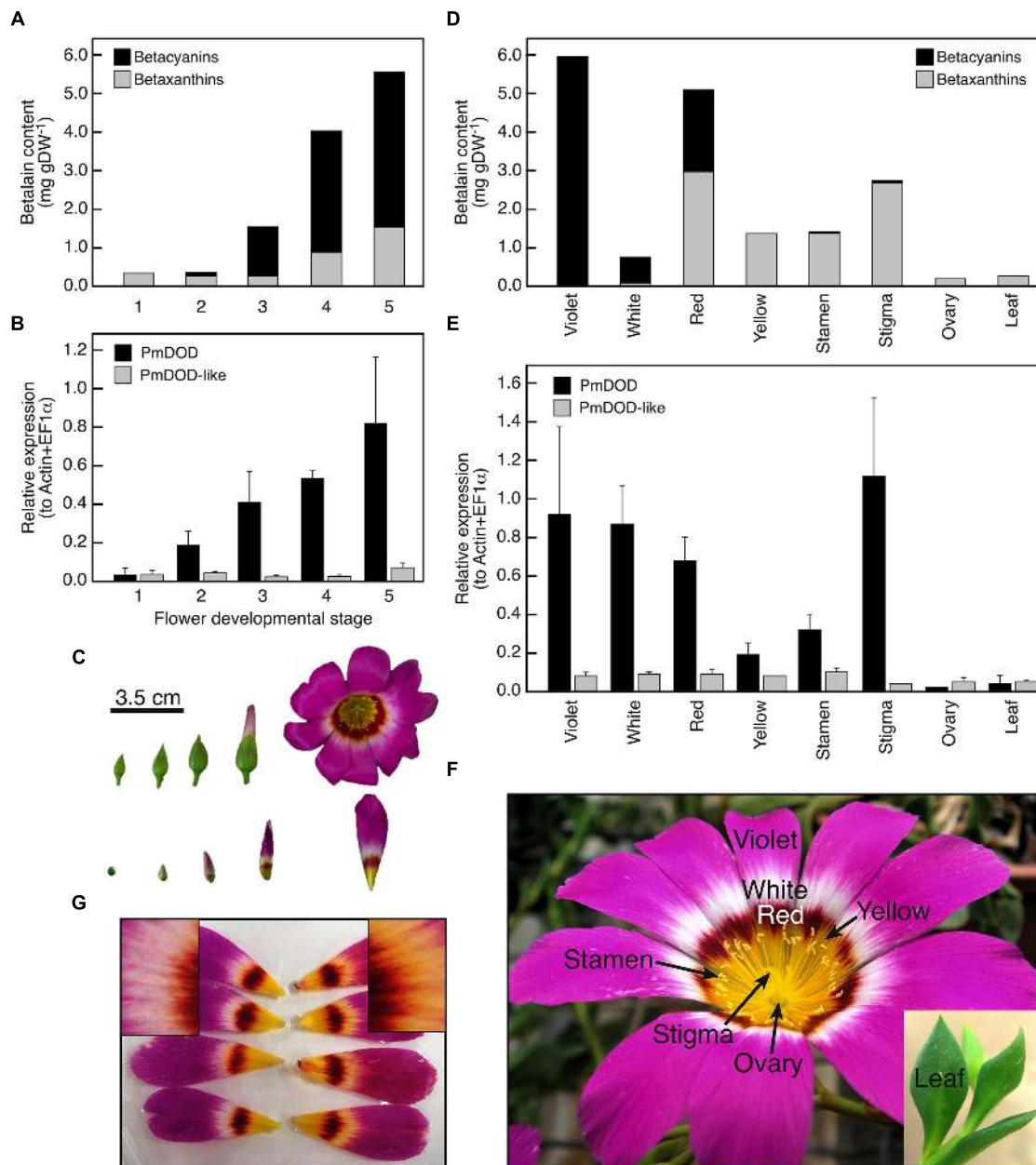
between stages 2 and 3, when petals are still enclosed within the sepals. A low amount of betaxanthins was present from stage 1, while betacyanins were first detected at stage 2. Ovaries and leaves, which are green tissues, had relatively low betaxanthin content and contained no detected betacyanins.

High-performance liquid chromatography profiles showed that the violet petal sectors of *P. mirabilis* contained four betacyanins, with no betaxanthin apparent (Figure 3). Peaks at 12.2, 13.6, 15.5, and 17.3 min retention times in the betacyanin gradient represented the glycosides betanin and iso-betanin, and the aglycones betanidin and iso-betanidin, respectively, as identified by LC-MS (Table 1). The aglycone betacyanin accounted for the large majority of the total betacyanin. The yellow petal sectors contained four different betaxanthins, the major one being vulgaxanthin I followed by dopaxanthin, vulgaxanthin II, and portulacaxanthin II. Red petal sectors contained the same betaxanthins in a similar proportion and only the two betacyanin aglycones, betanidin, and iso-betanidin (Figure 3). Due to the wide absorption region of betacyanins, they can also be discerned in the betaxanthin HPLC separation monitored at 470 nm, wherein they contribute ~30% of the total absorbance value at this wavelength (data not shown).

### Cloning and Expression Patterns of *Parakeelya* DOD and DOD-Like Sequences

To examine the genetic control of betalain biosynthesis in *P. mirabilis*, PCR was used to clone candidate LigB cDNAs from petal RNA. Using degenerate primers designed to Class II LigB sequences from other betalain species, cDNA fragments representing two different LigB genes were isolated. After RACE-PCR, full-length CDS was obtained for each. pBlast analysis using the NCBI non-redundant protein sequences database revealed that one showed highest identity to *Portulaca grandiflora* DODA (PgDODA; Q7XA48.1). As PgDODA has been shown to have a role in betalain production (Christinet et al., 2004) this gene was named DOD. *PmDOD* is 813 bp in length and encodes 271 amino acids. The other cDNA sequence had highest identity to *Phytolacca americana* DOD1 (PaDOD1; BAH66635.1). The contribution, if any, of PaDOD1 to betalain synthesis has not been resolved (Takahashi et al., 2009) therefore this *P. mirabilis* sequence was named DOD-like. *PmDOD-like* is 810 bp in length and encodes 270 amino acids. The *P. mirabilis* DOD and DOD-like amino acid sequences showed less identity to each other than to the heterologous sequences.

Spatio-temporal expression patterns of DOD and DOD-like in relation to betalain accumulation patterns in *P. mirabilis* were examined using quantitative PCR RNA analysis (Figure 2). During flower development, DOD transcript abundance correlated with the accumulation of betalain pigment while transcript abundance for DOD-like did not, and remained low. Transcript abundance was then compared across the different colored sectors within the petal as well as across other floral organs and leaves. Again, transcript abundance of DOD, but not DOD-like, generally correlated with betalain content (Figure 2). DOD-like transcript amounts were relatively low



**FIGURE 2 | Spatio-temporal patterns of betalain accumulation and DOD and DOD-like expression in *Parakeelya mirabilis*.** Accumulation (A) and expression (B) in petals during flower development (C), and accumulation (D) and expression (E) in different colored petal sectors (developmental stage 5) and other organs (F). Relative expression was determined by qPCR and

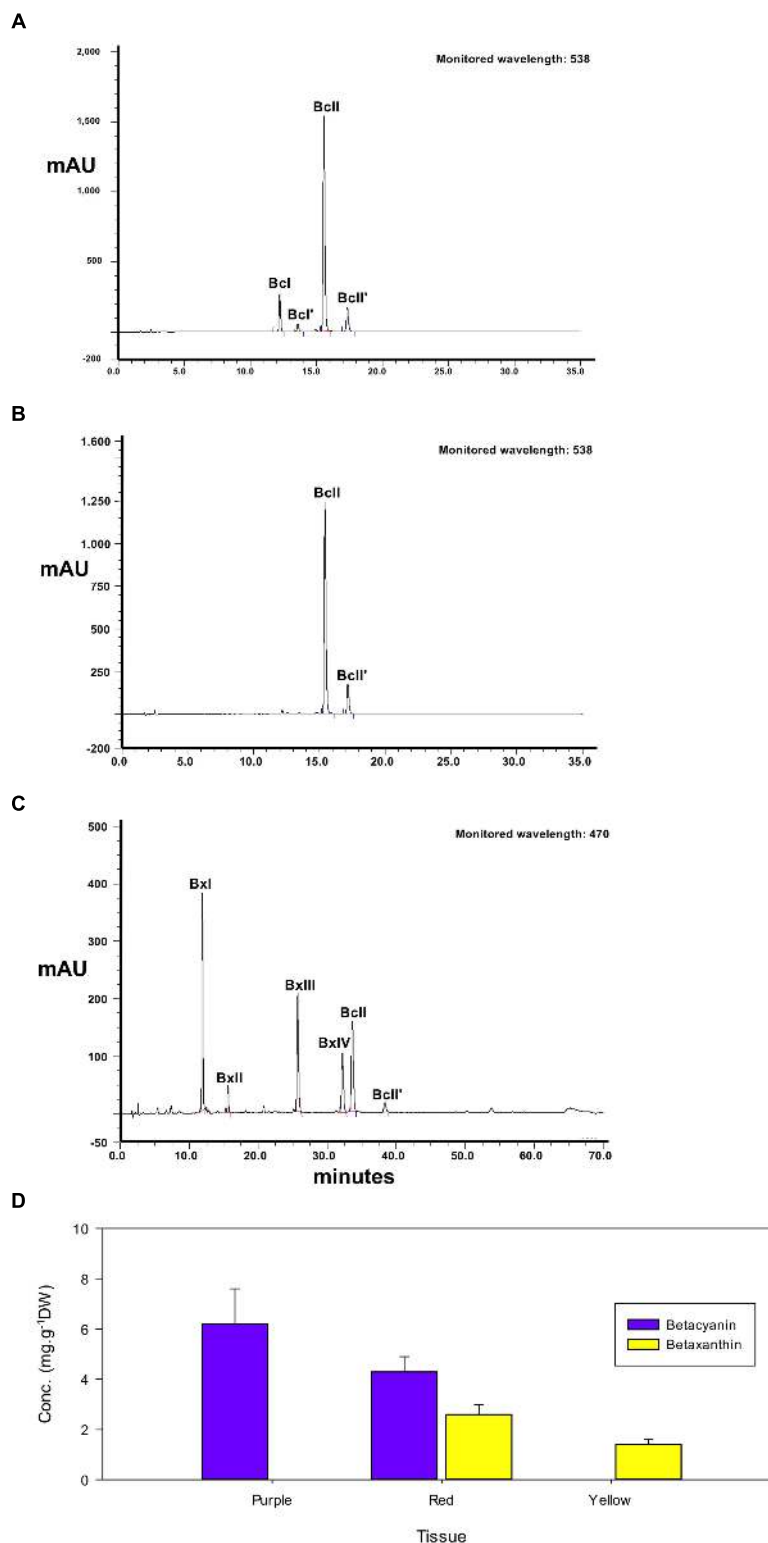
normalized to the geometric mean of *EF1α* and *ACT* expression. SE for three biological replicates is shown. (G) Shows petals fed a control solution (left) or a solution containing DOPA (right). The inserts show a magnification of the white petal region, with yellow pigmentation apparent in the DOPA but not control fed petals.

in all tissue samples. DOD was detected in the white sector at amounts similar to that for the violet and red sectors. This result was not an artifact of the sampling; the amount of contaminating pigmented tissue was only minor as indicated by the low amount of betalains detected (Figure 2), and DOPA feeding of the petal induced pigmentation in the white sector, indicating the presence of DOD activity (Figure 2). Overall, the results support a role for only DOD in betalain production,

with DOD-like having a different and as yet unidentified role.

### In planta Assays of PmDOD and PmDOD-like Activities

The potential for DOPA 4,5-extradiol cleavage activity by DOD and DOD-like was assayed using a biolistics-based transient gene expression assay (Harris et al., 2012). This



**FIGURE 3 | High-performance liquid chromatography elution profiles of betalain pigment extracts from petals of *Parakeelya mirabilis*.**

Extracts of violet (A) and red (B) petal sectors were separated by betacyanin elution gradient and monitored at 538 nm. Red (C) petal sectors were also separated by betaxanthin elution gradient and

monitored at 470 nm. The calculated proportion of total betacyanins or betaxanthins in each petal sector is shown in (D). Bcl, betanin; Bcl', iso-betanin; Bcll, betanidin; Bcll', iso-betanidin; Bxl, vulgaxanthin I; Bxll, vulgaxanthin II; Bxlll, dopaxanthin; BxlV, portulacaxanthin II. The peak IDs are identical to those in Table 1.

**TABLE 1 | Betalain pigment composition<sup>a</sup> in *Parakeelya mirabilis* petals.**

| Peak <sup>b</sup>              | Trivial name        |                   | Violet | Red  | Yellow | Rt <sup>e</sup> (min) | $\lambda_{\max}^f$ (nm) | m/z [M+H] <sup>+</sup> | m/z MS <sup>2</sup> |
|--------------------------------|---------------------|-------------------|--------|------|--------|-----------------------|-------------------------|------------------------|---------------------|
| <b>Betacyanin<sup>c</sup></b>  |                     | <b>Side chain</b> |        |      |        |                       |                         |                        |                     |
| Bcl                            | Betanin             | 5-O-glucoside     | 9.0    | -    | -      | 12.2                  | 534                     | 551                    | 389                 |
| Bcl'                           | Iso-betanin         | 5-O-glucoside     | 3.3    | -    | -      | 13.6                  | 533                     | 551                    | 389                 |
| Bcll                           | Betanidin           | Aglycone          | 68.7   | 77.0 | -      | 15.5                  | 540                     | 389                    | 343                 |
| Bcll'                          | Iso-betanidin       | Aglycone          | 19.0   | 23.0 | -      | 17.3                  | 540                     | 389                    | 343                 |
| <b>Betaxanthin<sup>d</sup></b> |                     | <b>Amino acid</b> |        |      |        |                       |                         |                        |                     |
| Bxl                            | Vulgaxanthin I      | Glutamine         | -      | 56.0 | 56.0   | 11.9                  | 469                     | 340                    | 323                 |
| Bxll                           | Vulgaxanthin II     | Glutamic acid     | -      | 5.0  | 7.0    | 15.6                  | 469                     | 341                    | 297                 |
| Bxlll                          | Dopaxanthin         | Dopa              | -      | 25.0 | 30.0   | 25.7                  | 470                     | 391                    | 347                 |
| BxlV                           | Portulacaxanthin II | Tyrosine          | -      | 14.0 | 7.0    | 32.2                  | 471                     | 375                    | 331                 |

<sup>a</sup>Proportion (%) of betacyanin and betaxanthin profiles as determined by peak integration.

-, not detectable.

<sup>b</sup>The peak IDs are identical to **Figure 3**.

<sup>c</sup>Separation in Betacyanin gradient monitored at 538 nm.

<sup>d</sup>Separation in Betaxanthin gradient monitored at 470 nm.

<sup>e</sup>Retention time.

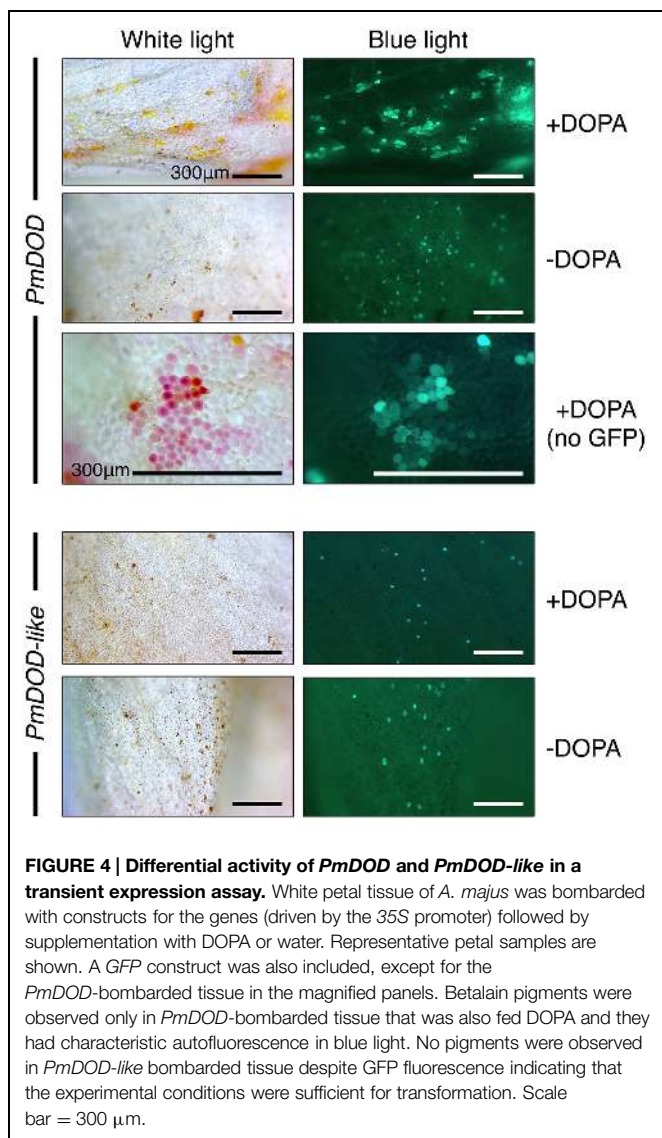
<sup>f</sup>Wavelength of maximum absorbance.

method uses the formation of betalain pigments as an indicator of 4,5-DOPA cleavage and betalamic acid (BA) formation. Petals of the white-flowered *nivea* (*chalcone synthase*) mutant of the anthocyanin-producing species *A. majus*, which lack anthocyanins and other flavonoids, were bombarded with the *P. mirabilis* cDNA constructs driven by the 35S promoter and fed DOPA. A construct with a *Green Fluorescent Protein* (GFP) reporter gene (35S:GFP) was co-precipitated with the other constructs onto the gold particles in some experiments as an indicator of transformation frequency. Pigment production was observed only in tissue bombarded with 35S:DOD and not in tissue bombarded with 35S:DOD-like or in control tissue (bombarded tissue fed water; **Figure 4**). The pigmented zones showed green autofluorescence under blue light in the absence of GFP, a characteristic typical of betaxanthins (Gandía-Herrero et al., 2005). The results for 35S:DOD were typical of what was observed in (Harris et al., 2012) when using a construct for a *Portulaca* DOD gene confirmed to be betalain-related. The pigments produced in the bombarded tissue were identified as the betacyanin betanin and the glutamine-based betaxanthin vulgaxanthin I based on HPLC comparison to the pigments in a beetroot standard sample (**Figure 5**). The pigment accumulated in multicellular zones, likely due to DOD protein or BA/pigment moving from the single transformed cells into adjacent untransformed cells. No pigment was observed in tissue bombarded with 35S:DOD-like despite GFP fluorescence showing the transformation process was successful (blue light panels, **Figure 4**). The DOD-like cDNA used in these assays was from a different accession of *P. mirabilis* and has two amino acid changes (ORF E192Q and T265S, both outside the proposed catalytic domain) compared to the original DOD-like sequence, which may represent allelic variation. It was confirmed that these amino acid changes were not influencing the assay by repeating the experiment with a construct using the original DOD-like sequence, with the same results obtained (Supplementary Figure S1).

When the transient assay was repeated for DOD-like without including the GFP internal control, a phenotype was revealed under blue-light. Single cells with weak green autofluorescence were observed at low frequency (**Figure 6A**). These cells were generally not visibly pigmented under white light, except for a few that possibly had a very pale pink color, suggesting extremely limited betalain production. No cells with these phenotypes were observed in any control tissue; viz. bombarded tissue fed water or tissue bombarded with empty vector or the GFP reporter construct and fed DOPA (data not shown). In an attempt to enhance any weak phenotypes, the GFP-free experiment was repeated using the higher concentration of 10 mM DOPA (rather than 1 mM). Close examination of the sporadic blue light fluorescent foci revealed that some foci had more white-light visible pigmentation than had been observed in the previous experiment (Supplementary Figure S1), albeit it was still very weak. Compared to DOD, which gave abundant and multicellular betalain-pigmented foci that were readily apparent even under much lower magnification (**Figure 4**, Supplementary Figure S2), both DOD-like constructs gave very limited activity. The low frequency of the cells precluded biochemical analysis of the phenotype from DOD-like, but these results suggest that DOD-like has, at best, only a very limited capacity for a DOPA cleavage that would allow BA formation.

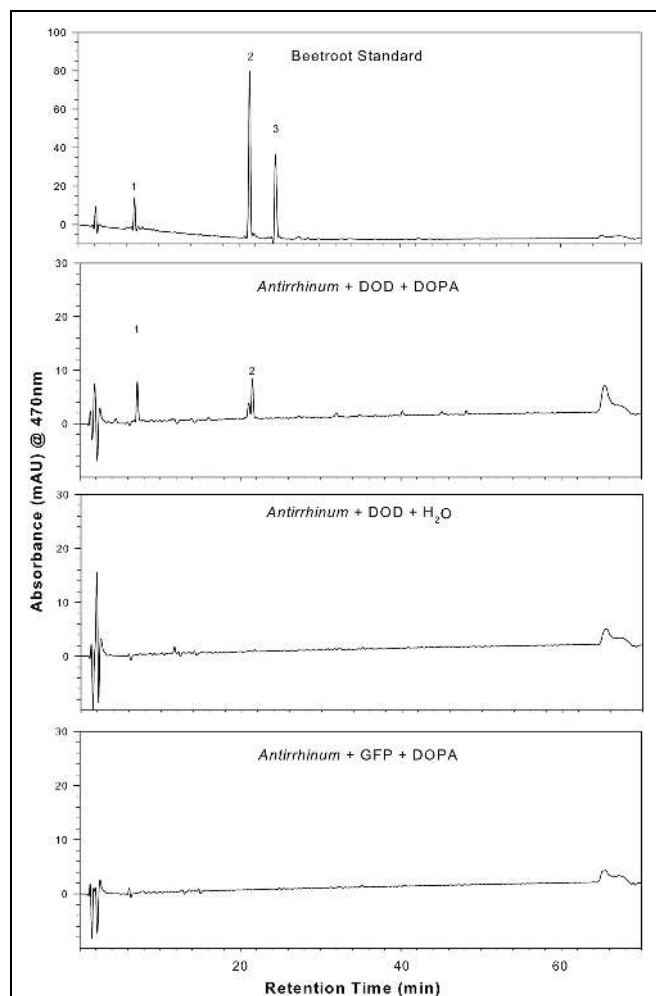
## The LigB Gene Family in *Parakeelya* and Other Caryophyllales

Given the differences in gene expression and protein function of the DOD and DOD-like cDNAs, the LigB gene family was examined in more detail for *P. mirabilis* and other Caryophyllales species. Two transcriptome databases for *P. mirabilis*, formed using 454 Next Generation Sequencing for RNA from either the violet (22,665 transcript contigs) or yellow (21,189 contigs) regions of the petal, were interrogated for sequences with similarity to LigB sequences. In addition to the previously identified DOD and DOD-like sequences, a third candidate LigB



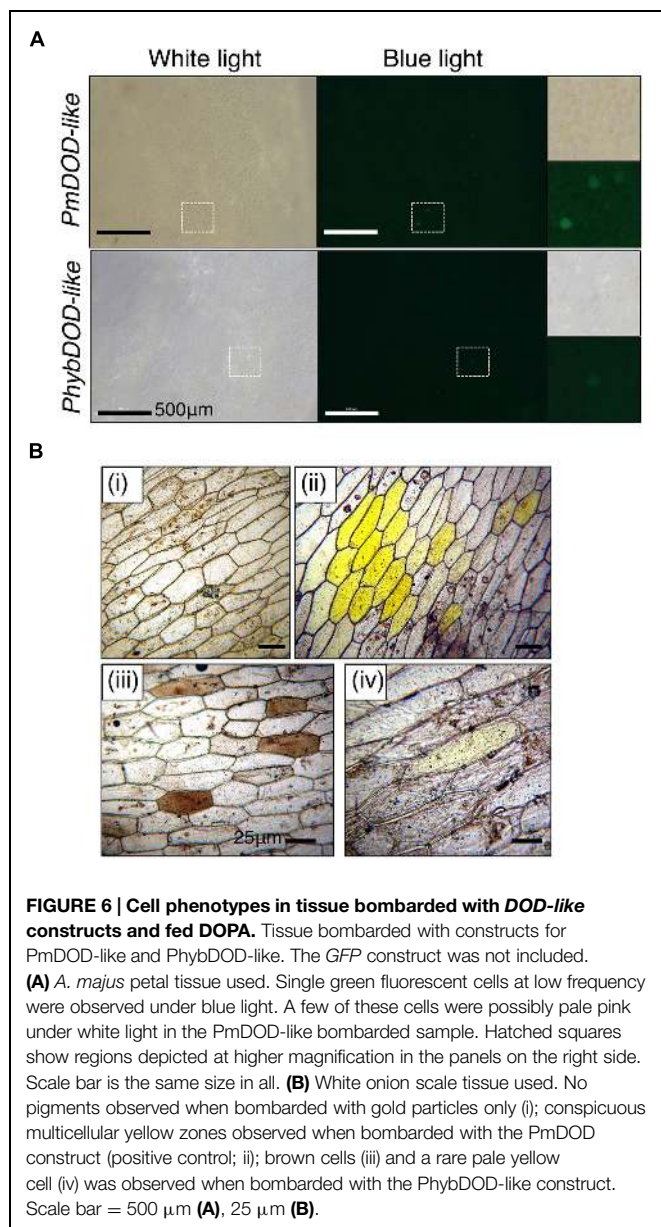
sequence was identified (named *PmLigB*) that contained the H-N-L-R motif characteristic of Class I LigB proteins (Christinet et al., 2004), **Figure 7**, and Supplementary Figure S3). This had 59.6 and 67.4% deduced amino acid identity to *PmDOD* and *PmDOD-like*, respectively. Given the sequence divergence this suggests that there are at least three functionally distinct *LigB* genes in *P. mirabilis*.

Publicly available sequence databases were interrogated to examine the occurrence of similar LigB gene families in non-Caryophyllales species and in betalain-producing or anthocyanin-producing species of the Caryophyllales. Examination of sequences from non-Caryophyllales species in NCBI/GenBank (<http://www.ncbi.nlm.nih.gov>) found only Class I type sequences, as based on the conserved region (**Figure 7** and data not shown). Examination of sequences from anthocyanin-producing species of the Caryophyllales in GenBank and a transcriptome database for the Caryophyllaceae species *Silene vulgaris* (Sloan et al., 2012), also found only



Class I type sequences (*Dianthus superbus* JL390264; *Silene vulgaris* isotig 21529). However, for the five betalain-producing families (11 species) of the Caryophyllales for which LigB sequences were available on NCBI, additional sequences were found that were more similar to *DOD* and *DOD-like* of *P. mirabilis* than they were to the Class I LigB (**Figure 7**). Including *P. mirabilis*, these represent the plant families of Amaranthaceae, Aizoaceae, Nyctaginaceae, Phytolaccaceae, and Portulacaceae.

The relationships within the LigB family were examined by forming phylogenetic trees on the DNA sequences (**Figure 8**). The resultant tree shows a clear separation of Class I LigB



sequences away from the additional LigB sequences found in betalain-producing species, with close to 100% bootstrap support. The additional LigB sequences form distinct clades consisting of those from the Amaranthaceae and those from the other betalain families examined, again with bootstrap support close to 100%. Within the non-Amaranthaceae clade, *PmDOD* and *PmDOD-like* are in separate groups with high bootstrap support, supporting the transient assay results that they have different biological activities. Within the Amaranthaceae species, the *Beta vulgaris* DODA1 separates from BvDODA, with BvDODA being close to *PhybDOD-like* (of *Ptilotus*), suggesting they too may have distinct activities. Trees formed on the amino acid sequences, with or without outgroup rooting, gave a very similar outcome to that from the DNA sequence (Supplementary Figure S4).

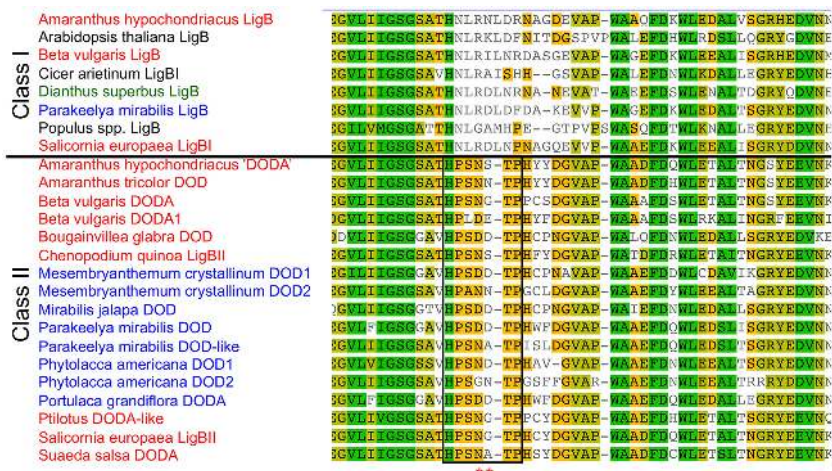
### In planta Assays of the Function of Amaranthaceae DOD and DOD-like Sequences

Given the differential biological activities and phylogenetic separation of *PmDOD* and *PmDOD-like*, the biological activity of Amaranthaceae genes with sequence similarity to *DOD* or *DOD-like* was examined. The characterized sequences from *B. vulgaris* (BvDODA and BvDODA1) were used along with a Class II homolog isolated from green leaf tissue of *Ptilotus* (*PhybDOD-like*), another Amaranthaceae species. Based on the amino acid sequences, BvDODA is closer to *PhybDOD-like* (80.2% identity) than it is to BvDODA1 (70.1% identity). Outside of the Amaranthaceae (Table 2), BvDODA is also more similar to *PmDOD-like* than *PmDOD* (or BvDODA1). The differences in % identities when examining BvDODA1 and *PmDOD* are less conclusive (Table 2), but in general they suggest that BvDODA, *PhybDOD-like*, and *PmDOD-like* may have a divergent enzymatic activity to that of BvDODA1 and *PmDOD*. Previous analysis has suggested BvDODA1 is the key LigB gene for betalain biosynthesis in *B. vulgaris* (Hatlestad et al., 2012), although BvDODA is also expressed in colored tissues (Hatlestad et al., 2012) and recombinant protein has some DOPA 4,5-cleavage activity (Sasaki et al., 2009; Gandía-Herrero and García-Carmona, 2012). The transient gene expression assay was used with constructs driven by the 35S promoter. Pigment production upon supplementation with DOPA was observed only in 35S:BvDODA1-bombarded tissue and not when 35S:BvDODA or 35S:*PhybDOD-like* was used (Figure 9). However, as seen with *PmDOD-like* (Figure 4), tissue bombarded with these other constructs had a low frequency of single cells with pale green autofluorescence under blue-light, but no visible pigmentation under white light (Figure 9). The transient expression assay was repeated for *PhybDOD-like* using *Allium cepa* (onion) epidermal cells, with *PmDOD* as a positive control. After DOPA feeding, yellow multicellular zones developed in *PmDOD* bombarded tissue, while sporadic brown cells and a single pale yellow cell developed in tissue bombarded with *PhybDOD-like* (Figure 6B). This demonstrated that the differential activity was not contingent on the host tissue used in the assay. These results provide further evidence of Class II LigB proteins in betalain producing taxa that are functionally different to DOD. These DOD-like enzymes, however, may conditionally exhibit limited DOPA cleavage activity that leads to BA.

A Class I LigB Caryophyllales representative (from *Dianthus superbis*, carnation) was also tested in the transient expression assay. No visible pigmentation or fluorescence under blue light was observed (Supplementary Figure S2), indicating that this class of protein does not have the same activity on DOPA in this assay as Class II types.

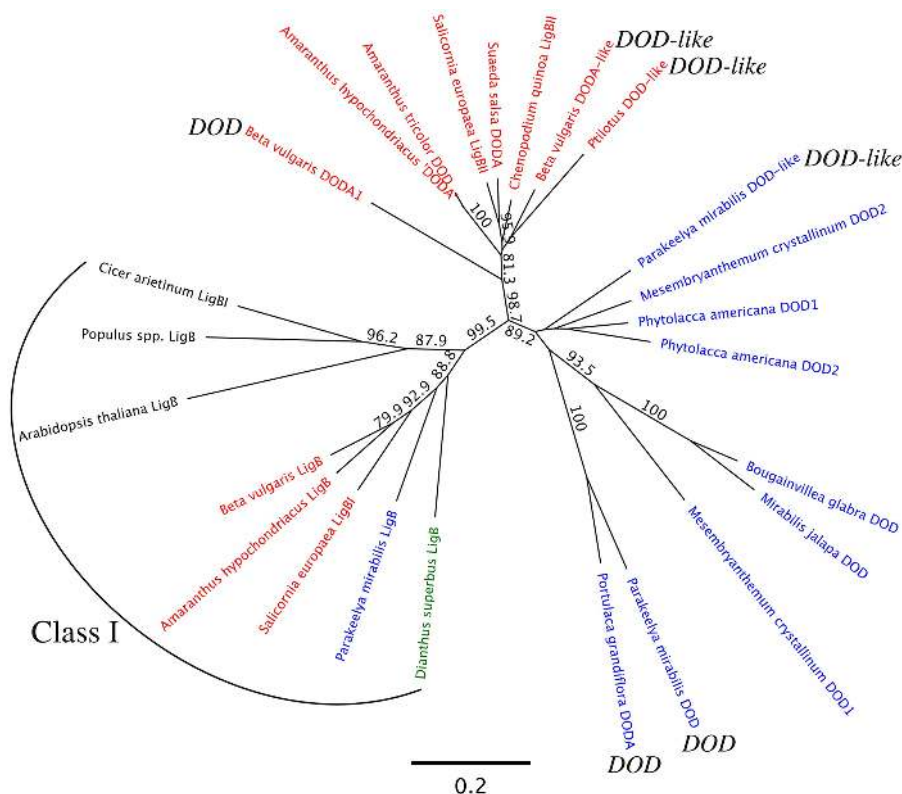
### The Genomic Structure of the LigB Family in Beta vulgaris

Diversification of gene function in plant secondary metabolism during evolution often involves localized gene duplication followed by neofunctionalisation, resulting in the tandem occurrence of related biosynthetic genes (Ober, 2005). The only betalain-producing species for which a genome sequence is



**FIGURE 7 | Sequence alignment of catalytic domain and adjacent regions of plant LigB homologs.** The sequences are shown divided into Classes I and II. The catalytic domain is boxed and the asterisks mark the acidic residues that may be important for DOD activity/DOPA recognition. For the Class II genes that have DOD in their name (aside from this study) the gene names used are those in the GenBank

annotations, which have generally been based only on sequence homology to DOD. The Amaranthaceae species names are in red, species of other betalain producing Caryophyllales families in blue, an anthocyanin taxa within the Caryophyllales (*Dianthus*) in green, and the non-Caryophyllales species in black. Accession numbers are given in Supplementary Figure S4.



**FIGURE 8 | Maximum likelihood analysis of LigB genes.** Analysis used the nucleotide sequences of the putative ORFs. The Amaranthaceae species names are in red, species of other betalain producing Caryophyllales families in blue, an anthocyanin taxa within the Caryophyllales (*Dianthus*) in green, and the non-Caryophyllales species in black. Sequences for McDOD1 and McDOD2 are not full-length but span the catalytic domain. Bootstrap analysis used 1000

datasets, with results shown for nodes that have at least 70% support. The branch-lengths indicate the average number of substitutions per site, with the scale bar given at the bottom of the figure. See Supplementary Figure S4 for the accession numbers. The Class I grouping is indicated. For the Class II sequences, the published DOD of *Portulaca grandiflora* is indicated along with the activities of sequences assayed in this study.

**TABLE 2 | Amino acid identity values (%) for deduced DOD and DOD-like sequences from *Beta vulgaris* and *Parakeelya mirabilis*, and DOD-like from *Ptilotus* sp. (Phyb), based on pairwise alignment.**

|               | BvDODA1 | BvDODA(-like) | PmDOD | PmDOD-like | PhybDOD-like |
|---------------|---------|---------------|-------|------------|--------------|
| BvDODA1       |         | 70.1          | 57.4  | 64.9       | 66.4         |
| BvDODA(-like) | 70.1    |               | 60.7  | 73.9       | 80.2         |
| PmDOD         | 57.4    | 60.7          |       | 70.4       | 60.4         |
| PmDOD-like    | 64.9    | 73.9          | 70.4  |            | 70.5         |
| PhybDOD-like  | 66.4    | 80.2          | 60.4  | 70.5       |              |

publically available is *B. vulgaris* (Dohm et al., 2014), so the structure and localisation in the genome of the three *LigB* sequences examined here was determined (Figure 10). All three genes have three exons, and the exon size and position of introns is conserved, with exon 1 being between 246 to 260 bp, exon 2 371 to 378 bp and exon 3 190 to 193 bp (Figure 10). *BvDODA1* is located on Chr2 while both the *ClassI LigB* and *DODA* are located on Chr4. The two genes on Chr4 are on the same genomic sequence scaffold, with the end of exon 3 of *DODA* being ~7.6 Kb upstream of exon 1 of the *ClassI LigB* gene (Figure 10).

### Characterization of the Catalytic Domain of DOD and DOD-like

The residues in the motif underlying our classification of the *LigB* genes are postulated to have a role in substrate recognition (Christinet et al., 2004). Given the differential activities in the transient assay it was examined whether there were conserved residues in the catalytic domain that may define a DOD protein (regardless of family taxon). Alignment of the sequences shows two adjacent acidic residues (D-D/E) in all confirmed DOD sequences that are also present in all sequences that are predicted to be DOD based on the phylogenetic analysis (Figure 8). To determine whether these residues are the reason of the differential activity observed in the transient assays, the predicted catalytic domain of PmDOD-like was mutagenised to convert HPSNATP to HPSDDTP, and the construct tested in the transient assay. The altered catalytic domain did not give activity that was characteristic of PmDOD (Supplementary Figure S2). However compared to the results with PmDOD-like, there were pigmented foci that were more readily detectable due to higher levels of pigmentation (Supplementary Figure S2).

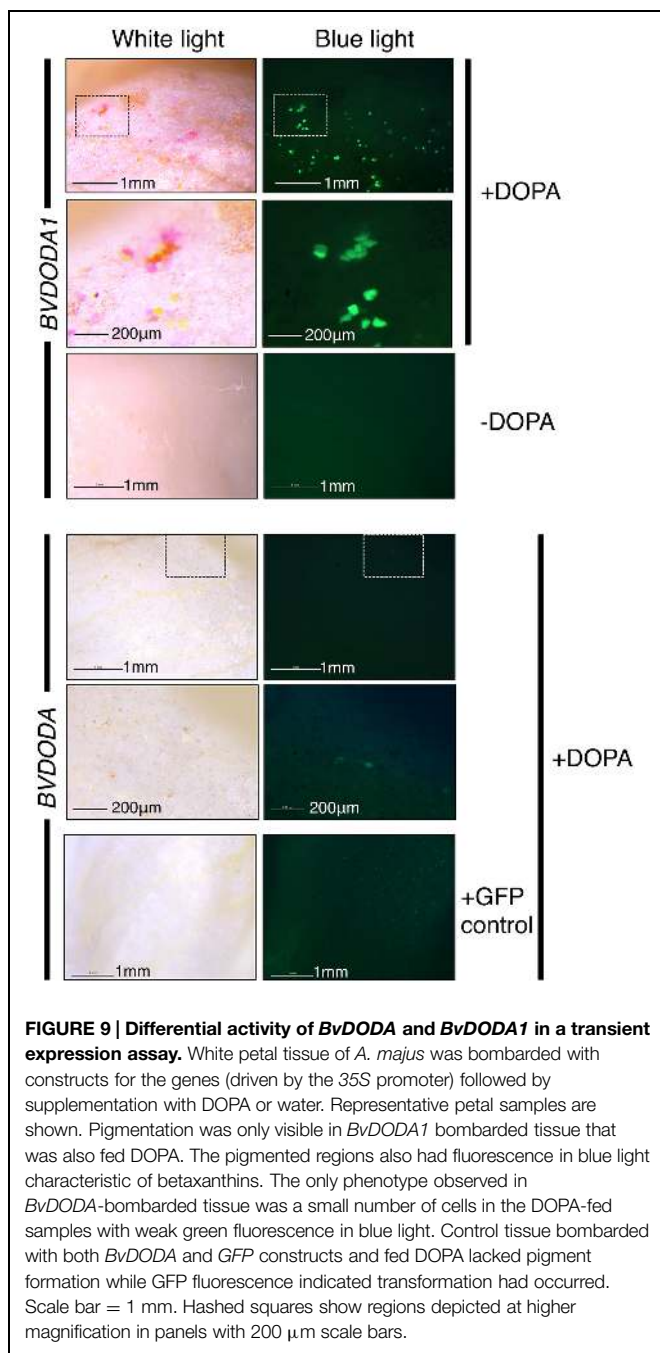
## Discussion

To gain further understanding of the betalain biosynthetic pathway, we have examined the pathway in *Parakeelya mirabilis* of the Montiaceae, which is phylogenetically well separated within the Caryophyllales from the established model of *Beta vulgaris*. Yellow betaxanthins were found at low amounts in leaves and ovaries and at higher amounts in all of the other floral organs examined. The red/purple betacyanins were only found at significant amounts in the petals, with only trace amounts present in the stamens and stigma. The petals contain both betacyanins and betaxanthins, with betacyanin aglycones

(betanidin and iso-betanidin) accounting for the large majority of the total betacyanin in the petals. Betacyanins, like anthocyanins, are typically glycosylated compounds. The role of glycosylation for betacyanins has not been established, but for anthocyanins it is essential for pigment stability and localisation into the vacuole (Tanaka et al., 2008; Cheynier et al., 2013). Although the accumulation of aglycone betacyanin is highly unusual, it has been previously reported in some species of the Portulacaceae and Aizoaceae, and in the Aizoaceae its presence was associated with violet flower colors while betanidin 5-O-glucoside (betanin) was associated with red flower colors (Strack et al., 1990). In *P. mirabilis* the area of betacyanins alone is also violet in color, while the red color is generated by the additional presence of yellow betaxanthins. Betacyanin aglycone occurrence in *P. mirabilis* is unlikely to be due to a lack of an appropriate glycosyltransferase gene, as some betanin and iso-betanin (which are glycosylated) was present.

The coloration of the *P. mirabilis* petals consists of violet, red, and yellow zones, with the violet and red separated by a colorless (white) zone. The violet petal sectors contained four betacyanins, with no betaxanthin apparent, while the yellow petal sectors contained four different betaxanthins, with no betacyanins apparent. The red sector color was the result of a mix of both betacyanins and betaxanthins. Also, the betaxanthins were present in the smallest flower buds examined, while betacyanins only appeared later in flower development. Thus, the flower has a series of distinct regulatory zones for betalain biosynthesis in which betacyanin and betaxanthin biosynthesis is separated both spatially and temporally. There has been no research on how such zones of betalain biosynthesis may be regulated at the molecular level. However, the formation of anthocyanin-based flower coloration patterns has been the focus of much research and R2R3MYB genes shown to be key to establishing such patterning in a range of species (Shang et al., 2011; Davies et al., 2012; Albert et al., 2014; Yamagishi et al., 2014; Yuan et al., 2014). As the only regulatory gene identified to date for betalain biosynthesis is an R2R3MYB that has proposed to be co-opted from anthocyanin regulation (Hatlestad et al., 2015), such genes would be excellent candidates to examine for the basis of the flower color patterning in *P. mirabilis* petals.

A cDNA for a protein with Dopa 4,5-dioxygenase (DOD) activity was identified from *P. mirabilis* and its transcript abundance examined across the same samples for which betalain analysis was conducted. The temporal pattern of DOD transcript during petal development closely reflected the pattern of betalain accumulation and also correlated to betalain amounts in the



**FIGURE 9 | Differential activity of *BvDODA* and *BvDODA1* in a transient expression assay.** White petal tissue of *A. majus* was bombarded with constructs for the genes (driven by the 35S promoter) followed by supplementation with DOPA or water. Representative petal samples are shown. Pigmentation was only visible in *BvDODA1* bombarded tissue that was also fed DOPA. The pigmented regions also had fluorescence in blue light characteristic of betaxanthins. The only phenotype observed in *BvDODA*-bombarded tissue was a small number of cells in the DOPA-fed samples with weak green fluorescence in blue light. Control tissue bombarded with both *BvDODA* and *GFP* constructs and fed DOPA lacked pigment formation while *GFP* fluorescence indicated transformation had occurred. Scale bar = 1 mm. Hashed squares show regions depicted at higher magnification in panels with 200 µm scale bars.

leaves, ovaries, stigmata and stamens. There are relatively few studies on whether transcript abundance for proteins with proven DOD activity correlates with betalain production. The clearest example is that *BvDODA1* transcript is found at high amounts in betalain producing tissues of *B. vulgaris* (Hatlestad et al., 2012), and the *BvDODA1* promoter is activated by the R2R3MYB betalain pathway activator (Hatlestad et al., 2015). The *P. mirabilis* data suggests regulation of the DOD gene expression is also a key aspect of regulation of the betalain biosynthetic pathway in this species. However, although the DOD transcript amounts correlated well with the overall amounts of betalains

in *P. mirabilis*, the only variation across the different colored sectors of the petal was that amounts were lower in the yellow sector. The DOD transcript amounts in the white sector were similar to that in the red and violet sectors, and DOPA feeding resulted in pigmentation, suggesting that the spatial regulation of betacyanin and betaxanthin production in these sectors operates at a different step than through transcriptional regulation of the DOD gene.

In addition to a cDNA for DOD, a second cDNA encoding a DOD-like protein was identified from *P. mirabilis*. The *DOD-like* transcript abundance did not correlate to betalain levels, either in different tissues or across flower development (Figure 2), suggesting it has a role unrelated to betalain biosynthesis. Examination of public databases identified sequences with high % identity to either *DOD* or *DOD-like* from a range of Caryophyllales species. These 'Class II' LigB sequences showed clear separation from the 'Class I' type on amino acid alignments (Figure 7) and phylogenetic trees formed from either the nucleotide or deduced amino acid sequences (Figure 8 and Supplementary Figure S4). Within the Class II, the Amaranthaceae species separated from representatives from four other Caryophyllales families. In the non-Amaranthaceae grouping of these sequences there was a separation that was not on a species basis, with groupings containing the *DOD* and *DOD-like* sequences whose function was assayed in this study. There were insufficient numbers of sequences to determine if this was also the case for the Amaranthaceae species, but *BvDODA1* did separate from *BvDODA*. Given this suggestion of a larger LigB gene family consistently occurring in betalain producing plants, and the transcript abundance of *PmDOD-like* not reflecting betalain occurrence in *Parakeelya*, the function of *PmDOD-like* and *DOD-like* sequences from two Amaranthaceae species (*B. vulgaris* and *Ptilotus*) was examined using the transient assay system. None of the three sequences conferred betalain production in the way that *DOD* sequences did (Figures 4, 6, and 9), further suggesting that the *DOD-like* ones encode a distinct enzyme activity. What the activity (or activities) might be is unknown, but the *DOD-like* transgenes did confer production of weakly fluorescent and pigmented cells. Thus, it may be an activity closely related to that of DOD. In this regard, it should be noted that while in our assay the *DOD-like* *BvDODA* did not demonstrate equivalent activity to confirmed DOD proteins, BA formation *in vitro* using recombinant *BvDODA* protein produced in *Escherichia coli* has been reported (Sasaki et al., 2009; Gandía-Herrero and García-Carmona, 2012) and there is a preliminary report of DOPA 4,5-cleavage activity *in vitro* for recombinant versions of some non-Caryophyllales plant LigB genes (presumably Class I types; Sasaki et al., 2007). However, it is difficult to know how reflective of *in planta* activity these *in vitro* assays are, as no comparative data has been presented for a DOD protein that has a confirmed role in betalain synthesis.

Although a distinct region of the peptide can distinguish Class I and Class II sequences, no single conserved motif was apparent that could be used to further separate sequences within Class II that might account for the different activities observed with DOD from *DOD-like* (Figure 7 and Supplementary

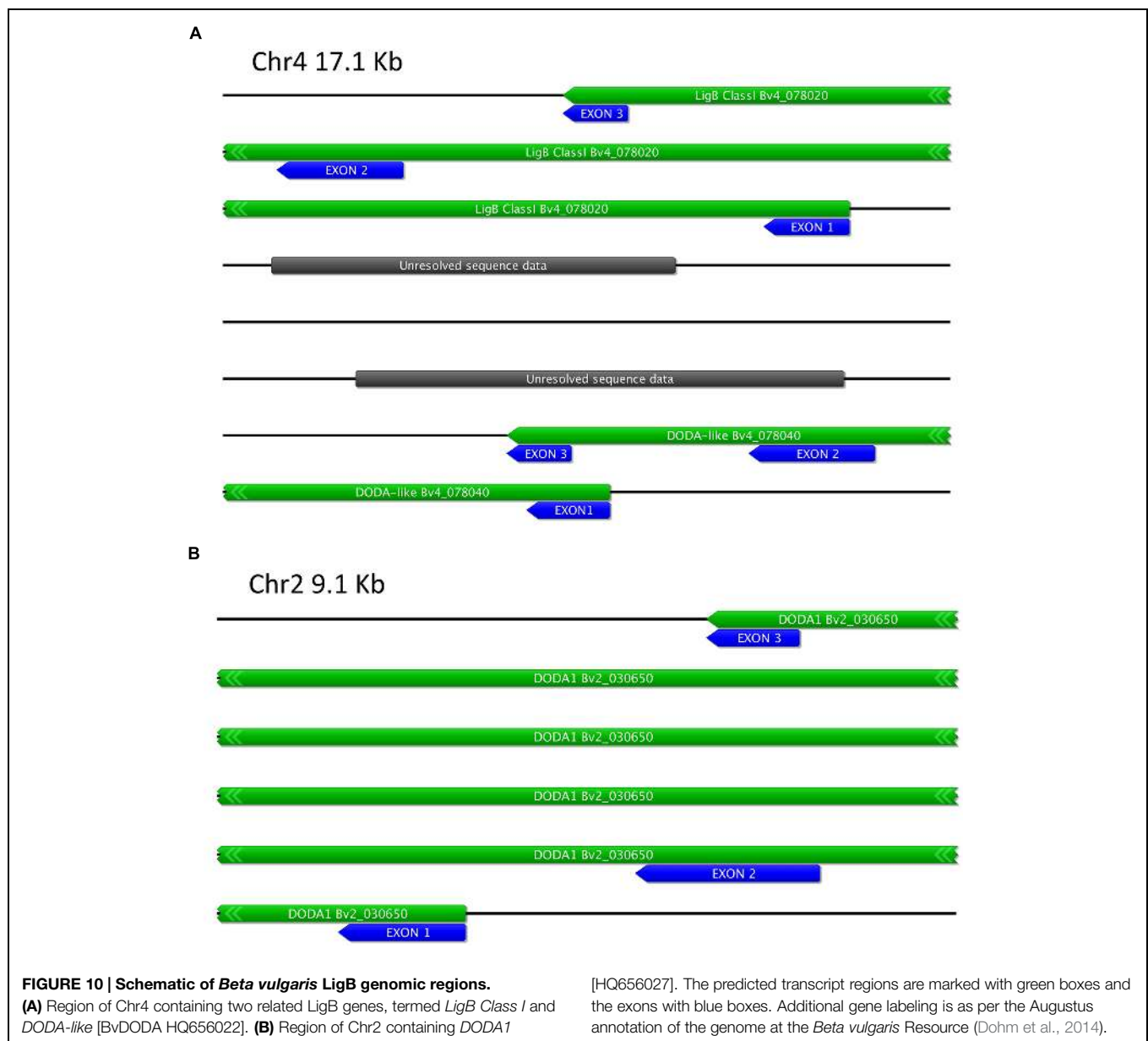


Figure S3). However, a comparison of the region postulated to be involved in substrate recognition (Christinet et al., 2004) revealed two adjacent acidic residues that may be important to DOD. These are found in PgDODA, PmDOD, and BvDODA1 (Figure 7), all of which gave clear betalain production in the transient expression assay (this paper and Harris et al., 2012). However, the lack of DOD-equivalent activity in the transient assay by PmDOD-like that had been mutated to gain these acidic residues found in DOD showed that on their own they were not sufficient to convert a DOD-like protein to DOD. Therefore, all the key residues that differentiate DOD from DOD-like are yet to be defined. Although the substituted residues did not generate DOD-equivalent activity, there was a visual increase in the weak pigmentation that could be produced in transformed cells, indicating that the residues may make some

contribution to enzymatic efficiency and would be worthy of further investigation.

Published studies on the transcript abundance patterns for some of the DOD-like sequences from other species (as based on our alignments) also suggest they may not be involved in betalain biosynthesis and could be a part explanation of why 'DOD' transcript levels often do not always correlate with betalain production. The stress induced *DODA1* and *DODA2* of *Amaranthus hypochondriacus* (Casique-Arroyo et al., 2014) likely encode *DODA-like* and *LigB Class I* sequences, respectively. The transcript abundance pattern of these genes did not consistently match the pattern of betalain production with regard to drought or herbivory stress, an observation that also led Casique-Arroyo et al. (2014) to suggest *AhDODA2* may have a role outside of betalain biosynthesis. Similarly, the stress induced

*DODA* of *Suaeda salsa* is also likely a *DODA-like* sequence (Zhao et al., 2010, 2011), and its transcript abundance also only correlated with betalain production in some of the studies.

Only Class I *LigB* genes were found in the databases for taxa outside the Caryophyllales while both classes of *LigB* genes were found in a range of Caryophyllales species. One possibility is that a Class I gene duplication event gave rise to the Class II gene lineage, as this is a common event in plant secondary metabolism gene evolution (Ober, 2005). Such duplication often results in the tandem occurrence of related biosynthetic genes, and this was found to be the case when examining the *B. vulgaris* genome (the only betalain species for which a genome sequence is publically available). The conserved exon/intron structure of the three *B. vulgaris* genes examined here and the co-localization of the Class I *LigB* and *DOD-like* genes (Figure 10) supported a common gene ancestry. Investigation of the gene structure in other species would be of much interest to help answer the questions on the *LigB* gene family evolution.

## Author Contributions

H-HC, HN, BM, KC, and KS conducted experiments; KS, KD, and RC conducted sequence analyses; DL contributed to experimental analyses; KS, DH, KD, DJ contributed to project conception, experimental design, data interpretation; KG, KS,

and KD supervised KC; DH and DJ supervised H-HC and BM; KS, DH, KD, and H-HC prepared the manuscript; All authors read and approved the final manuscript.

## Acknowledgments

For H-HC, DH, BM, and DJ, funding for the project was facilitated with seed funding from a UQ Early Career Research Grant awarded to DH from The University of Queensland, followed by additional project funding from the Australian Research Council (ARC) Linkage Projects scheme (Australian Government; Project LP0883314) in collaboration with Aussie Colours Pty. Ltd., and the Centre for Native Floriculture (CNF) based at the University of Queensland, Gatton Campus. DL, HN, KC, KD, and KS were supported by the New Zealand Ministry for Business, Innovation and Employment contract C02X0805. We also thank Cathie Martin for the *nivea* line, Steve Arathoon for technical assistance in the extraction and quantification of the betalain pigments, and Ian King and Julie Ryan for the care of the plants at Palmerston North.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00499>

## References

- Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., et al. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* 26, 962–980. doi: 10.1105/tpc.113.122069
- Bahramnejad, B., Erickson, L. R., and Goodwin, P. H. (2010). Induction of expression and increased susceptibility due to silencing a 4,5-DOPA dioxygenase extradiol-like gene of *Nicotiana benthamiana* in the interaction with the hemibiotrophic pathogens, *Colletotrichum destructivum*, *Colletotrichum orbiculare* or *Pseudomonas syringae* pv. tabaci. *Plant Sci.* 178, 147–157. doi: 10.1016/j.plantsci.2009.10.007
- Brockington, S., Alexandre, R., Ramdial, J., Moore, M., Crawley, S., Dhirga, A., et al. (2009). Phylogeny of the Caryophyllales sensu lato: revisiting hypotheses on pollination biology and perianth differentiation in the core Caryophyllales. *Int. J. Plant Sci.* 170, 627–643. doi: 10.1086/597785
- Brockington, S. F., Walker, R. H., Glover, B. J., Soltis, P. S., and Soltis, D. E. (2011). Complex pigment evolution in the Caryophyllales. *New Phytol.* 190, 854–864. doi: 10.1111/j.1469-8137.2011.03687.x
- Carolin, R. C. (1993). "Portulacaceae," in: *The Families and Genera of Vascular Plants*, eds K. Kubitzki, V. Bittrich, and J. Rohwer (Berlin: Springer), 544–545. doi: 10.1007/978-3-662-02899-5\_64
- Casique-Arroyo, G., Martínez-Gallardo, N., González de la Vara, L., and Delano-Frier, J. P. (2014). Betacyanin biosynthetic genes and enzymes are differentially induced by (a)biotic stress in *Amaranthus hypochondriacus*. *PLoS ONE* 9:e99012. doi: 10.1371/journal.pone.0099012
- Chevieux, B., Wetter, T., and Suhai, S. (1999). "Genome sequence assembly using trace signals and additional sequence information," in *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, Vol. 99, Hanover, 45–56.
- Cheyrier, V., Comte, G., Davies, K. M., Lattanzio, V., and Martens, S. (2013). Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Biochem.* 72, 1–20. doi: 10.1016/j.plaphy.2013.05.009
- Christinet, L., Burdet, X., Zaiko, M., Hinz, U., and Zryd, J. (2004). Characterization and functional identification of a novel plant 4,5-extradiol dioxygenase involved in betalain pigment biosynthesis in *Portulaca grandiflora*. *Plant Physiol.* 134, 265–274. doi: 10.1104/pp.103.031914
- Clement, J., Mabry, T., Wyler, H., and Dreiding, A. (1994). "Chemical review and evolutionary significance of the betalains," in: *Caryophyllales: Evolution and Systematics*, eds H.-D. Behnke, and T. Mabry (Berlin: Springer), 247–261. doi: 10.1007/978-3-642-78220-6\_11
- Cuenoud, P., Savolainen, V., Chatrou, L. W., Powell, M., Grayer, R. J., and Chase, M. W. (2002). Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *Am. J. Bot.* 89, 132–144. doi: 10.3732/ajb.89.1.132
- Davies, K. M. (2015). Swapping one red pigment for another. *Nat. Genet.* 47, 5–6. doi: 10.1038/ng.3174
- Davies, K. M., Albert, N. W., and Schwinn, K. E. (2012). From landing lights to mimicry: the molecular regulation of flower colouration and mechanisms for pigment patterning. *Funct. Plant Biol.* 39, 619–638. doi: 10.1071/fp12195
- Dohm, J. C., Minoche, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., et al. (2014). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505, 546–549. doi: 10.1038/nature12817
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Fischer, N., and Dreiding, A. S. (1972). Biosynthesis of betalains. On the cleavage of the aromatic ring during the enzymatic transformation of Dopa into betalamic acid. *Helvetica Chimica Acta* 55, 649–658. doi: 10.1002/hlca.19720550240
- Gandía-Herrero, F., and García-Carmona, F. (2012). Characterization of recombinant *Beta vulgaris* 4,5-DOPA-extradiol-dioxygenase active in the biosynthesis of betalains. *Planta* 236, 91–100. doi: 10.1007/s00425-012-1593-2
- Gandía-Herrero, F., García-Carmona, F., and Escribano, J. (2005). Fluorescent pigments: new perspectives in betalain research and applications. *Food Res. Int.* 38, 879–884. doi: 10.1016/j.foodres.2005.01.012
- Giannasi, D. E. (1988). "Flavonoids and evolution in the dicotyledons," in: *The Flavonoids: Advances in Research Since 1980*, ed. J. B. Harborne (London: Chapman and Hall), 479–504. doi: 10.1007/978-1-4899-2913-6\_14
- Gleave, A. P. (1992). A versatile binary vector system with a T-DNA organizational structure conducive to efficient integration of cloned DNA

- into the plant genome. *Plant Mol. Biol.* 20, 1203–1207. doi: 10.1007/BF00028910
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520
- Harris, N. N., Javellana, J., Davies, K. M., Lewis, D. H., Jameson, P. E., Derolles, S. C., et al. (2012). Betalain production is possible in anthocyanin-producing plant species given the presence of DOPA-dioxygenase and L-DOPA. *BMC Plant Biol.* 12:34. doi: 10.1186/1471-2229-12-34
- Hatlestad, G. J., Akhavan, N. A., Sunnadaniya, R. M., Elam, L., Cargile, S., Hembd, A., et al. (2015). The beet Y locus encodes an anthocyanin MYB-like protein that activates the betalain red pigment pathway. *Nat. Genet.* 47, 92–96. doi: 10.1038/ng.3163
- Hatlestad, G. J., Sunnadaniya, R. M., Akhavan, N. A., Gonzalez, A., Goldman, I. L., McGrath, J. M., et al. (2012). The beet R locus encodes a new cytochrome P450 required for red betalain production. *Nat. Genet.* 44, 816–820. doi: 10.1038/ng.2297
- Hershkovitz, M. A. (1998). *Parakeelya*: a new genus segregated from *Calandrinia* (Portulacaceae). *Phytologia* 84, 98–106.
- Hinz, U. G., Fivaz, J., Girod, P. A., and Zrýd, J. P. (1997). The gene coding for the DOPA dioxygenase involved in betalain biosynthesis in *Amanita muscaria* and its regulation. *Mol. Gen. Genet.* 256, 1–6. doi: 10.1007/s004380050539
- Kimler, L., Mears, J., Mabry, T. J., and Rösler, H. (1970). On the question of the mutual exclusiveness of betalains and anthocyanins. *Taxon* 19, 875–878. doi: 10.2307/1218301
- Kugler, F., Stintzing, F. C., and Carle, R. (2004). Identification of betalains from petioles of differently colored Swiss Chard (*Beta vulgaris* L. ssp. *cicla* [L.] Alef. Cv. Bright Lights) by high-performance liquid chromatography-electrospray ionization mass spectrometry. *J. Agric. Food Chem.* 52, 2975–2981. doi: 10.1021/jf035491w
- Kugler, F., Stintzing, F. C., and Carle, R. (2007). Characterisation of betalain patterns of differently coloured inflorescences from *Gomphrena globosa* L. and *Bougainvillea* sp. by HPLC-DAD-ESI-MSn. *Anal. Bioanal. Chem.* 387, 637–648. doi: 10.1007/s00216-006-0897-0
- Mabry, T. J., Taylor, A., and Turner, B. L. (1963). The betacyanins and their distribution. *Phytochemistry* 2, 61–64. doi: 10.1016/S0031-9422(00)88016-4
- Nyffeler, R., and Eggli, U. (2010). Disintegrating portulacaceae: a new familial classification of the suborder Portulacineae (Caryophyllales) based on molecular and morphological data. *Taxon* 59, 227–240.
- Obbens, F. J. (2006). A review of the tuberous *Calandrinia* species (section *Tuberosae*), including three new species for Western Australia. *Nuytsia* 16, 95–115.
- Ober, D. (2005). Seeing double: gene duplication and diversification in plant secondary metabolism. *Trends Plant Sci.* 10, 444–449. doi: 10.1016/j.tplants.2005.07.007
- Pfaffl, M. W., Tichopad, A., Prgommet, C., and Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: bestKeeper - Excel-based tool using pair-wise correlation. *Biotechnol. Lett.* 26, 509–515. doi: 10.1023/B:BILE.0000019559.84305.47
- Sakuta, M. (2014). Diversity in plant red pigments: anthocyanins and betacyanins. *Plant Biotechnol. Rep.* 8, 37–48. doi: 10.1007/s11816-013-0294-z
- Sasaki, N., Abe, H., Goda, H., Adachi, T., Kasahara, K., and 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 Cell Physiol.* 50, 1012–1016. doi: 10.1093/pcp/pcp053
- Sasaki, N., Nagira, Y., Yabuki, K., Koda, T., Adachi, T., and Ozeki, Y. (2007). Isolation of DOPA 4,5-dioxygenase homologs from betalain non-producing plants. *Plant Cell Physiol.* 48, S64.
- Schliemann, W., Kobayashi, N., and Strack, D. (1999). The decisive step in betaxanthin biosynthesis is a spontaneous reaction. *Plant Physiol.* 119, 1217–1232. doi: 10.1104/pp.119.4.1217
- Schliemann, W., Steiner, U., and Strack, D. (1998). Betanidin formation from dihydroxyphenylalanine in a model system. *Phytochemistry* 49, 1593–1598. doi: 10.1016/S0031-9422(98)00276-3
- Shang, Y. J., Schwinn, K. E., Bennett, M. J., Hunter, D. A., Waugh, T. L., Pathirana, N. N., et al. (2007). Methods for transient assay of gene function in floral tissues. *Plant Methods* 3, 1. doi: 10.1186/1746-4811-3-1
- Shang, Y. J., Venail, J., Mackay, S., Bailey, P. C., Schwinn, K. E., Jameson, P. E., et al. (2011). The molecular basis for venation patterning of pigmentation and its effect on pollinator attraction in flowers of *Antirrhinum*. *New Phytol.* 189, 602–615. doi: 10.1111/j.1469-8137.2010.03498.x
- Sloan, D. B., Keller, S. R., Berardi, A. E., Sanderson, B. J., Karpovich, J. F., and Taylor, D. R. (2012). *De novo* transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae). *Mol. Ecol. Res.* 12, 333–343. doi: 10.1111/j.1755-0998.2011.03079.x
- Strack, D., Marxen, N., Reznik, H., and Ihlenfeld, H.-D. (1990). Distribution of betacyanins and hydroxycinnamic acid-glucose esters in flowers of the *Ruschieae*. *Phytochem.* 29, 2175–2178. doi: 10.1016/0031-9422(90)83032-V
- Strack, D., Vogt, T., and Schliemann, W. (2003). Recent advances in betalain research. *Phytochemistry* 62, 247–269. doi: 10.1016/S0031-9422(02)00564-2
- Suzuki, M., Miyahara, T., Tokumoto, H., Hakamatsuka, T., Goda, Y., Ozeki, Y., et al. (2014). Transposon-mediated mutation of CYP76AD3 affects betalain synthesis and produces variegated flowers in four o'clock (*Mirabilis jalapa*). *J. Plant Physiol.* 171, 1586–1590. doi: 10.1016/j.jplph.2014.07.010
- Svenson, J., Smallfield, B. M., Joyce, N. I., Sansom, C. E., and Perry, N. B. (2008). Betalains in red and yellow varieties of the andean tuber crop ulluco (*Ullucus tuberosus*). *J. Agric. Food Chem.* 56, 7730–7737. doi: 10.1021/jf8012053
- Takahashi, K., Takamura, E., and Sakuta, M. (2009). Isolation and expression analysis of two DOPA dioxygenases in *Phytolacca americana*. *Z. Naturforsch. C* 64, 564–573. doi: 10.1515/znc-2009-7-816
- Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-313X.2008.03447.x
- Wang, C.-Q., Song, H., Gong, X.-Z., Hu, Q.-G., Liu, F., and Wang, B.-S. (2007). Correlation of tyrosinase activity and betacyanin biosynthesis induced by dark in C-3 halophyte *Suaeda salsa* seedlings. *Plant Sci.* 173, 487–494. doi: 10.1016/j.plantsci.2007.07.010
- Weng, J.-K., Li, Y., Mo, H., and Chapple, C. (2012). Assembly of an evolutionarily new pathway for  $\alpha$ -pyrone biosynthesis in *Arabidopsis*. *Science* 337, 960–964. doi: 10.1126/science.1221614
- Wickramasinghe, P., Harrison, D. K., and Johnston, M. E. (2009). Reproductive biology and intergeneric breeding compatibility of ornamental *Portulaca* and *Calandrinia* (Portulacaceae). *Austr. J. Bot.* 57, 697–707. doi: 10.1071/BT09107
- Yamagishi, M., Toda, S., and Tadaki, K. (2014). The novel allele of the *LhMYB12* gene is involved in splatter-type spot formation on the flower tepals of Asiatic hybrid lilies (*Lilium* spp.). *New Phytol.* 201, 1009–1020. doi: 10.1111/nph.12572
- Yuan, Y. W., Sagawa, J. M., Frost, L., Vela, J. P., and Bradshaw, H. D. Jr. (2014). Transcriptional control of floral anthocyanin pigmentation in monkeyflowers (*Mimulus*). *New Phytol.* 204, 1013–1027. doi: 10.1111/nph.12968
- Zhao, S.-Z., Sun, H.-Z., Chen, M., and Wang, B.-S. (2010). Light-regulated betacyanin accumulation in euhalophyte *Suaeda salsa* calli. *Plant Cell Tissue Organ Cult.* 102, 99–107. doi: 10.1007/s11240-010-9710-z
- Zhao, S.-Z., Sun, H.-Z., Gao, Y., Sui, N., and Wang, B.-S. (2011). Growth regulator-induced betacyanin accumulation and dopa-4,5-dioxygenase (DODA) gene expression in euhalophyte *Suaeda salsa* calli. *In Vitro Cell. Dev. Biol. Plant* 47, 391–398. doi: 10.1007/s11627-011-9339-6

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Citation:** Chung H-H, Schwinn KE, Ngo HM, Lewis DH, Massey B, Calcott KE, Crowhurst J, Joyce DC, Gould KS, Davies KM and Harrison DK (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. *Front. Plant Sci.* 6:499. doi: 10.3389/fpls.2015.00499

Copyright © 2015 Chung, Schwinn, Ngo, Lewis, Massey, Calcott, Crowhurst, Joyce, Gould, Davies and Harrison. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.