

RESEARCH ARTICLE

Open Access



Hybrid de novo transcriptome assembly of poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) bracts

Vinicius Vilperte^{1,2†} , Calin Rares Lucaci^{3†}, Heidi Halbwirth⁴, Robert Boehm², Thomas Rattei^{3*} and Thomas Debener^{1*}

Abstract

Background: Poinsettia is a popular and important ornamental crop, mostly during the Christmas season. Its bract coloration ranges from pink/red to creamy/white shades. Despite its ornamental value, there is a lack of knowledge about the genetics and molecular biology of poinsettia, especially on the mechanisms of color formation. We performed an RNA-Seq analysis in order to shed light on the transcriptome of poinsettia bracts. Moreover, we analyzed the transcriptome differences of red- and white-bracted poinsettia varieties during bract development and coloration. For the assembly of a bract transcriptome, two paired-end cDNA libraries from a red and white poinsettia pair were sequenced with the Illumina technology, and one library from a red-bracted variety was used for PacBio sequencing. Both short and long reads were assembled using a hybrid de novo strategy. Samples of red- and white-bracted poinsettias were sequenced and comparatively analyzed in three color developmental stages in order to understand the mechanisms of color formation and accumulation in the species.

Results: The final transcriptome contains 288,524 contigs, with 33% showing confident protein annotation against the TAIR10 database. The BUSCO pipeline, which is based on near-universal orthologous gene groups, was applied to assess the transcriptome completeness. From a total of 1440 BUSCO groups searched, 77% were categorized as complete (41% as single-copy and 36% as duplicated), 10% as fragmented and 13% as missing BUSCOs. The gene expression comparison between red and white varieties of poinsettia showed a differential regulation of the flavonoid biosynthesis pathway only at particular stages of bract development. An initial impairment of the flavonoid pathway early in the color accumulation process for the white poinsettia variety was observed, but these differences were no longer present in the subsequent stages of bract development. Nonetheless, *GSTF11* and *UGT79B10* showed a lower expression in the last stage of bract development for the white variety and, therefore, are potential candidates for further studies on poinsettia coloration.

Conclusions: In summary, this transcriptome analysis provides a valuable foundation for further studies on poinsettia, such as plant breeding and genetics, and highlights crucial information on the molecular mechanism of color formation.

Keywords: Poinsettia (*Euphorbia pulcherrima*), RNA-Seq, Anthocyanin, Hybrid de novo transcriptome, Bract coloration

* Correspondence: thomas.rattei@univie.ac.at; debener@genetik.uni-hannover.de

[†]Vinicius Vilperte and Calin Rares Lucaci contributed equally to this work.

³Department of Microbiology and Ecosystem Science, University of Vienna, 1090 Vienna, Austria

¹Institute of Plant Genetics, Leibniz Universität Hannover, 30419 Hannover, Germany

Full list of author information is available at the end of the article



Background

The poinsettia, *Euphorbia pulcherrima* Willd. ex Klotsch, also known as Nochebuena or Christmas Star, is one of the most important ornamental potted plants around the globe. The species is native to Mexico [76] and belongs to the family Euphorbiaceae and genus *Euphorbia*, with the latest estimate containing around 2000 species and representing one of the largest genera within angiosperms [31]. The species is known by its red bract coloration, which is due to the accumulation of anthocyanin pigments. Anthocyanins are a class of flavonoid secondary metabolite compounds [48] which provide orange to blue colors to flowers, seeds, fruits and other vegetative tissues in plants [72]. Moreover, they have multiple functional roles in plant-environment interactions, such as light protection and antioxidants, chelating agents for metals [43], as well as protection against biotic and abiotic stresses [2, 19]. The molecular mechanism involved in anthocyanin biosynthesis has been extensively described for several species [59], but only scarce information is currently available for poinsettia [30, 57].

In ornamental poinsettia, there is a coexistence of green, reddish, and red leaves/bracts [54] in the same plant, which implies a constant regulation of the anthocyanin and adjacent pathways throughout the bract development process. A bract is a modified or specialized leaf, often associated with a reproductive structure such as a flower or inflorescence. In poinsettia, bract axillary buds differentiate into flowers [36] under short day conditions, which is accompanied by the development and coloration of bracts, thus indicating that the anthocyanin metabolism is regulated by photoperiodism [34]. The color range in poinsettia varieties is obtained either through classical breeding (crossing) or mutagenic breeding (radiation), thus generating a spectrum of bract colors, such as pink, marble (pink center surrounded by white margins) and white/creamy. The pink coloration in pink and marble bracts are due to periclinal chimeric structures [55], while the reason for white/creamy coloration remains uncertain. Since the expression of all structural genes and the related enzyme activities involved in the formation of red anthocyanin pigments can be determined, the appearance of acyanic (uncolored) varieties is here referred to as the '*white paradox*'. The elucidation of such mechanisms is extremely valuable for this crop since the production of plants with bright and/or different colors is a key aspect for breeding and consumer acceptance [30]. Despite the popularity of poinsettia, information about its genome and transcriptome have not been generated yet. Transcriptome assemblies are very useful in elucidating the major transcripts and isoforms involved in pigmentation pathways, as well as their expression profiles under specific conditions [3, 24, 47, 96].

De novo transcriptome assemblies still represent a challenge for non-model plant species, where the general approach relies on the use of short cDNA sequences (such as Illumina technology). Some of the issues faced are related to the sensitivity of alignment errors due to paralogs and multigene families, production of artefactual chimeras and fragmented genes, and potentially misestimated allelic diversity [17]. The recent use of PacBio technology has generated an improvement in various plant transcriptomes [5, 80, 87] since it is able to generate full-length transcripts without the need of assembly algorithms. Nevertheless, long reads generated by the PacBio technology show an error rate of 13–15% [6] and, therefore, deep sequencing is required to correct the errors based on base coverage. As an alternative, a hybrid assembly approach (combining short and long reads) could be implemented to achieve similar results. Although still scarce, some methods have shown the applicability and usefulness of this approach to improve transcriptome annotations [25, 56, 84].

With the aim of generating valuable information on molecular aspects of poinsettia, we have assembled and functionally annotated a de novo bract transcriptome for the species. In addition, we also underlined and characterized the regulation of the main pathways involved in the transition of green leaves to colored bracts. Lastly, we characterized the main differences between red- and white-bracted poinsettia varieties, focusing on the flavonoid and adjacent pathways that are involved in pigment accumulation in plant tissues. Due to tissue-specific expression and the difficulty of recovering low expressed transcripts, the de novo assembled transcriptome is not expected to represent the entire range of transcripts of the species; nevertheless, the successful assembly of different isoforms and the differential expression analysis enabled a first insight into the *white paradox*.

Results

De novo assembly and functional annotation of the poinsettia bract transcriptome

In order to create a representative transcriptome for poinsettia bracts, cDNA libraries of the variety pair Christmas Feelings (red) and Christmas Feelings Pearl (white) were sequenced using the Illumina NextSeq500 system. In addition, a full-length cDNA library, from the Vintage variety (red), was sequenced using the PacBio Sequel System. After quality control and data cleaning, 36,989,889 and 35,404,728 Illumina reads were generated for the red and white varieties, respectively, with an average proportion of 77.4% clean reads for the libraries. The Iso-Seq pipeline v3.0 was applied to the PacBio dataset and, after sequence classification, clustering, and quality control, a total of 30,768 high-quality full-length transcripts were generated (Table 1).

Table 1 Summary of Illumina and PacBio sequencing

Illumina sequencing			
Variety	Total number of reads	Remained reads after rRNA removal	Remained reads after quality trimming (QV ≥ 20)
Christmas Feelings	46,734,786	43,267,294	36,989,889
Christmas Feelings Pearl	46,772,696	42,704,780	35,404,728
PacBio sequencing			
Variety	Total number of CCS	Number of FLNC reads	Number of polished transcripts
Vintage	72,202	52,077	30,768

We mapped the Illumina post-processed reads to the PacBio transcripts to assess their completeness and to verify if they represent a significant portion of the transcriptome. The distribution of average coverage over the full-length transcripts is shown in Additional file 1. The majority of the full-length transcripts were covered by both Illumina datasets. Out of 30,768 full-length transcripts, 1987 were not covered by the Illumina reads from the red variety, while 1808 were not covered by the reads from the white variety. Moreover, the overall mapping rate was 60 and 58% of read pairs for the red and white varieties, respectively. These results imply that the PacBio transcripts did not seem to capture the majority of the bract transcriptome of poinsettia, thus not suitable to be used as the only dataset for our transcriptome. To overcome that, a hybrid de novo assembly strategy was applied.

The Trinity tool was used to perform the de novo assembly with both Illumina and PacBio post-processed reads. The final assembly contains 288,524 contigs belonging to 138,702 genes, with a total of 257,619,354 assembled bases, GC content of 38.23% and an N50 of 1488. To evaluate the quality and coverage of the assembled transcripts, the Illumina reads were re-mapped to the final transcriptome using bowtie2. The re-mapping ratio was 83 and 81% for Christmas Feelings and Christmas Feelings Pearl, respectively. Next, the assembled transcripts were annotated against TAIR10 and SwissProt databases. From 288,524 total contigs assembled, 78,350 (27.1%) showed annotation against the SwissProt database, while 95,900 (33.2%) of them showed homology to *A. thaliana* transcripts (TAIR10), both using an *E-value* < 1E-20. Due to the higher number of retrieved annotations, we used the data from TAIR10 for further analyses. A total of 14,623 *A. thaliana* homologous transcripts were identified in our transcriptome (Additional file 2), with 6105 showing a length coverage between 90 and 100% (Additional file 3). Functional annotation and Gene Ontology (GO) terms

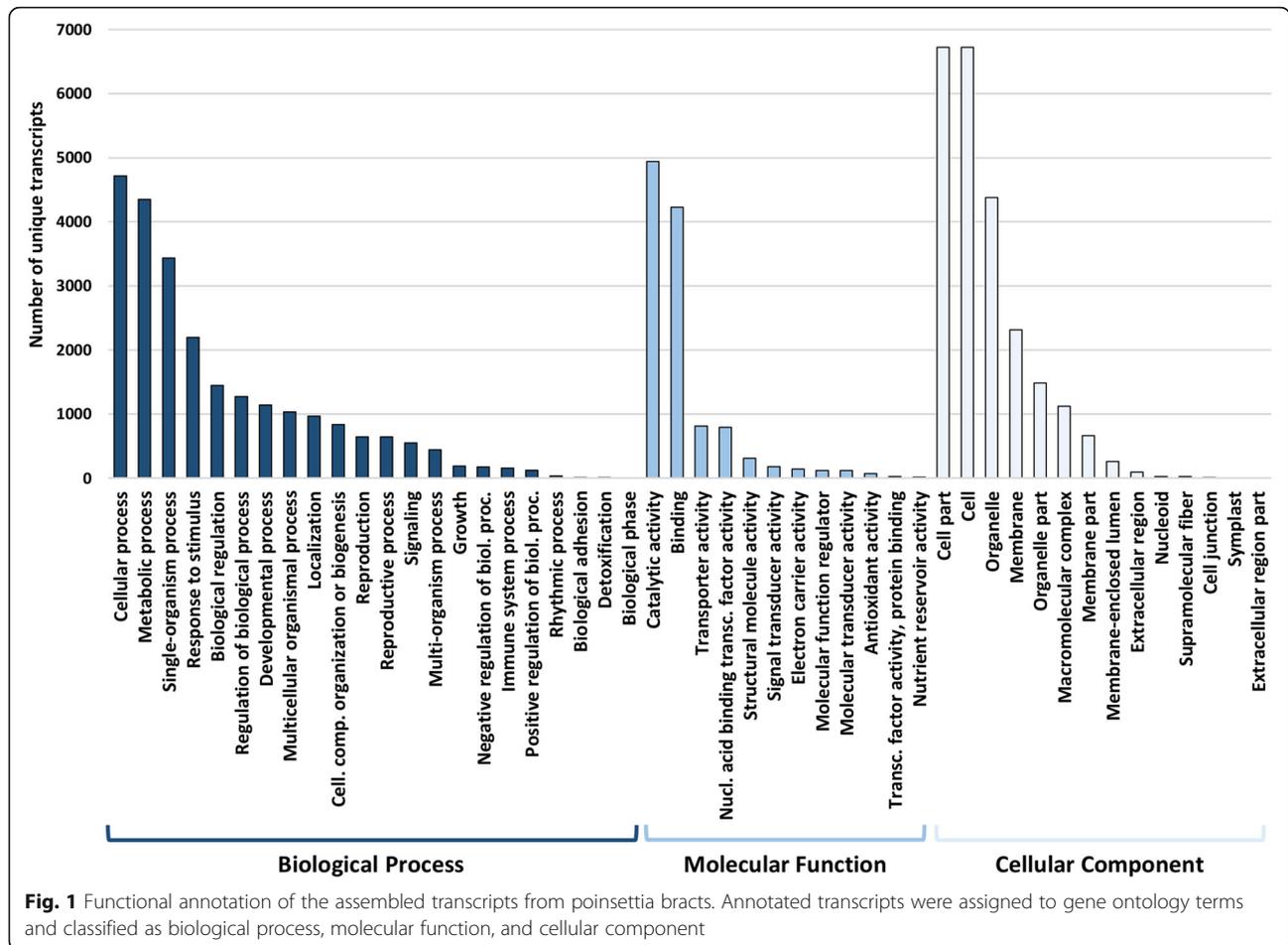
were retrieved using the online tool agriGO. Out of the 14,623 different *A. thaliana* homologous transcripts, 13,809 (94.4%) were assigned to one or more GO terms. On the other hand, 814 homologous transcripts (representing 6261 transcripts in our transcriptome) could not be assigned to GO terms.

In total, 13,809 unique transcripts were functionally characterized in 48 subcategories and grouped in three main groups: biological process (22 subcategories), molecular function (12) and cellular component (14), with several transcripts annotated with multiple GO terms (Fig. 1). Within the biological process category, cellular process (4716) and metabolic process (4348) were prominent, indicating a higher number of genes involved in important metabolic activities. In the molecular function category, the majority of the GO terms were grouped into catalytic activity (4941) and binding (4225), followed by transporter (811) and nucleic acid binding (791) activities. For the cellular component category, 6721 GO terms were assigned to both cell and cell part, and, together with organelle (4376) and membrane (2314), represent the dominant transcripts in this category.

Several genes related to the flavonoid biosynthetic pathway were identified in our bract transcriptome. The annotation against the TAIR10 database revealed 127 transcripts belonging to 23 known flavonoid-related structural genes and 24 transcripts belonging to six flavonoid-related transcription factors (Table 2). The genes with the highest number of identified transcripts were *Flavone 3'-O-methyltransferase 1* (15), *Hydroxycinnamoyl-CoA shikimate transferase* (12) and *Dihydroflavonol 4-reductase* (11). On the other hand, *Phenylalanine ammonia-lyase 4*, *Flavanone 3-hydroxylase* and *TTG1 Transducin/WD40 repeat-like* were the only genes that contained a single transcript. Similar genes were identified in another poinsettia transcriptome, also with a high number of transcripts assigned to different genes [30]. Moreover, it is important to note that, due to the lack of an available genome, poinsettia specific transcripts might not have been identified and, therefore, a higher number of transcripts might be involved in the flavonoid pathway. The expression of several flavonoid-related genes found in our transcriptome, as well as previous metabolite profiling studies [30, 68], implies that poinsettia bract pigmentation is achieved through the regulation of those genes and further accumulation of flavonoid compounds.

Transcriptome completeness and comparison to related species

A transcriptome represents the complete set and quantity of transcripts from a specific stage of development or physiological condition [78]. By relying on bract material to assemble the transcriptome of poinsettia, transcripts specific to other plant tissues, e.g. root and stem,



could be missing in bracts. For a better overview of the completeness of the poinsettia bract transcriptome generated in the present study, publicly available sequences from root, stem and leaf tissues of *Euphorbia pekinensis* were retrieved and individual transcriptomes for each tissue were assembled and annotated. Based on the annotation against the TAIR10 database, tissue-specific transcripts were observed for each of the *E. pekinensis* transcriptomes. A total of 2149 Arabidopsis homologous proteins from all three *E. pekinensis* transcriptomes were not present in our poinsettia bract transcriptome. From these proteins, 317 were uniquely present in the leaf transcriptome, while 346 and 235 homologous proteins were uniquely detected in root and stem transcriptomes, respectively. On the other hand, 1262 Arabidopsis homologous proteins present on the bract transcriptome were not detected in any of the *E. pekinensis* transcriptomes.

The BUSCO pipeline, which is based on near-universal orthologous gene groups, was applied to assess the completeness of the newly assembled poinsettia bract transcriptome, as well as the *E. pekinensis* transcriptomes. This pipeline permits to assess the completeness of

transcriptomes based on evolutionarily informed expectations of gene content. Therefore, it enables like-for-like quality comparisons of different data sets (e.g. transcriptomes) [83]. From a total of 1440 BUSCO (embryophyta_odb9 database) groups searched, the poinsettia bract transcriptome showed 1115 (77%) categorized as complete (595 (41%) as single-copy and 520 (36%) as duplicated), 139 (10%) as fragmented and 186 (13%) as missing BUSCOs (Table 3). The BUSCO results for the *E. pekinensis* transcriptomes are also shown in Table 3.

When comparing the completeness of the poinsettia bract with the tissue-specific transcriptomes from *E. pekinensis*, we noticed that the number of complete BUSCOs is comparable in all transcriptomes, but with poinsettia showing a lower percentage of duplicated ones. Additionally, the number of fragmented and missing BUSCOs also showed similar percentages. Out of 186 missing BUSCOs in the bract transcriptome (12.9%), 136 of them were identified in at least one of the *E. pekinensis* transcriptomes, with 16 exclusively present in the leaf transcriptome and another 16 exclusively present in the root transcriptome. The most abundant orthologs among those groups belonged to the Pentatricopeptide

Table 2 List of flavonoid biosynthesis related genes identified in the poinsettia bract transcriptome

	<i>A. thaliana</i> orthologous	Gene	Enzyme name	# of transcripts identified	
Structural genes	AT2G37040.1	<i>PAL1</i>	Phenylalanine ammonia-lyase 1	4	
	AT3G53260.1	<i>PAL2</i>	Phenylalanine ammonia-lyase 2	5	
	AT3G10340.1	<i>PAL4</i>	Phenylalanine ammonia-lyase 4	1	
	AT2G30490.1	<i>C4H</i>	Trans-cinnamate-4-hydroxylase	8	
	AT1G51680.1	<i>4CL1</i>	4-coumarate: CoA ligase 1	3	
	AT3G21240.1	<i>4CL2</i>	4-coumarate: CoA ligase 2	5	
	AT1G65060.1	<i>4CL3</i>	4-coumarate: CoA ligase 3	4	
	AT3G21230.1	<i>4CL5</i>	4-coumarate: CoA ligase 5	2	
	AT5G13930.1	<i>CHS</i>	Chalcone synthase	4	
	AT5G48930.1	<i>HCT</i>	Hydroxycinnamoyl-CoA shikimate transferase	12	
	AT5G05270.1	<i>CHI3</i>	Chalcone-flavonone isomerase 3	2	
	AT3G55120.1	<i>CHI1</i>	Chalcone-flavonone isomerase 1	4	
	AT3G51240.1	<i>F3H, FHT</i>	Flavanone 3-hydroxylase	1	
	AT5G07990.1	<i>F3'H</i>	Flavonoid 3'-hydroxylase	10	
	AT5G08640.1	<i>FLS1</i>	Flavonol synthase 1	5	
	AT5G42800.1	<i>DFR</i>	Dihydroflavonol 4-reductase	11	
	AT4G22880.1	<i>LDOX, ANS</i>	Leucoanthocyanidin dioxygenase, Anthocyanin synthase	3	
	AT5G17050.1	<i>UGT78D2</i>	UDP-glucosyl transferase 78D2	10	
	AT3G29590.1	<i>5MAT</i>	Anthocyanidin 5-O-glucoside-6"-O-malonyltransferase	2	
	AT1G61720.1	<i>ANR</i>	Anthocyanidin reductase	10	
	AT5G54160.1	<i>OMT1</i>	Flavone 3'-O-methyltransferase 1	15	
	AT3G59030.1	<i>TT12</i>	TRANSPARENT TESTA 12	5	
	AT3G03190.1	<i>GSTF11</i>	Glutathione S-transferase	4	
	Transcription factors	AT1G63650.1	<i>EGL1</i>	Transcription factor EGL1	7
		AT1G66370.1	<i>MYB113</i>	MYB domain protein 113	2
		AT2G47460.1	<i>MYB12</i>	MYB domain protein 12	4
AT5G35550.1		<i>TT2</i>	Transcription factor TT2 (MYB123)	6	
AT4G09820.1		<i>TT8</i>	Transcription factor TT8 (bHLH42)	4	
AT5G24520.1		<i>TTG1</i>	TRANSPARENT TESTA GLABRA 1	1	

repeat (PPR) superfamily protein. In addition, 50 ortholog groups are equally missing in all four transcriptomes, with the majority of them also belonging to PPR superfamily protein groups. On the other hand, 171 ortholog groups present in the bract transcriptome were completely absent from all three *E. pekinensis* transcriptome. The list of missing BUSCO orthologs for one or more of

the transcriptomes is available in Additional file 4. All in all, the BUSCO analysis shows that tissue-specific orthologs might be absent in our poinsettia bract transcriptome. Nevertheless, a high level of transcriptome completeness was observed and thus enables us to reliably use the data for further analyses.

Table 3 Completeness assessment of *E. pulcherrima* and *E. pekinensis* transcriptomes by the BUSCO pipeline

Species - Tissue	Complete BUSCOs		Fragmented BUSCOs	Missing BUSCOs
	Single-copy	Duplicated		
<i>E. pulcherrima</i> - Bract	41.3%	36.1%	9.7%	12.9%
<i>E. pekinensis</i> - Leaf	31.3%	50.1%	8.9%	9.7%
<i>E. pekinensis</i> - Root	32.0%	46.5%	10.2%	11.3%
<i>E. pekinensis</i> - Stem	36.3%	41.0%	9.1%	13.6%

Differential expression analysis of poinsettia bracts

To understand the dynamics of gene expression in different stages of bract and color development of poinsettia, RNA-Seq libraries from three independent biological replicates of the Christmas Feelings and Christmas Feelings Pearl varieties, sampled at three developmental stages (Stage 1 - S1, Stage 2 - S2 and Stage 3 - S3), were sequenced for transcriptome analysis. In total, 927,560,033 million raw reads with a length of 75 bp were obtained and, after quality trimming and rRNA removal,

an average of 91.6% reads remained available. The overall mapping of the datasets against the poinsettia bract transcriptome was 92.9% (Additional file 5). In addition, a high correlation between biological replicates (Pearson correlation) was observed, thus showing the reliability of the datasets (Additional file 6).

The RNA-Seq data from the three bract developmental stages were compared using two different approaches. First, we aimed to characterize the variation in gene expression between the different stages of bract development, regardless of the bract color. Hereof, we compared the six samples from S1 (three Christmas Feelings and three Christmas Feelings Pearl as independent biological replicates) against the six samples from S2, as well as S2 against S3. Secondly, we were interested in analyzing the differences between red and white bracts for each of the time points, especially those related to biosynthesis and accumulation of pigments. To this end, we compared the Christmas Feelings and Christmas Feelings Pearl varieties of each stage against each other.

Characterization of the expression profiles of poinsettia bracts during three developmental stages

To characterize the gene regulation dynamics in the transition of green leaves to fully developed bracts, six independent biological replicates (three replicates from Christmas Feelings and three replicates from Christmas Feelings Pearl) for three bract developmental stages were analyzed. The pairwise comparison for the first transition point, between S1 and S2, showed significantly lower expression rates for 3743 transcripts in S2. A pathway enrichment analysis of the DEGs was performed and 39 GO terms were differentially enriched (False Discovery Rate (FDR) ≤ 0.05). The enriched pathways linked to major biological processes included: i) response to temperature stimulus (GO:0009266); ii) enzyme-linked receptor protein signaling pathway (GO:0007167); and iii) response to heat (GO:0009408). On the other hand, 2675 transcripts were higher expressed in the S2 samples. Pathway enrichment analysis showed that 22 GO terms were differentially enriched, with the major molecular functions enriched pathways being related to: i) catalytic activity (GO:0003824); ii) oxidoreductase activity (GO:0016491); and iii) peptidase activity (GO:0008233).

For the second transition point, S2 to S3, 4479 transcripts had significantly lower expression in S3. A total of 104 GO terms were differentially enriched, with the major biological processes being related to response to temperature stimulus (GO:0009266) and photosynthesis (GO:0015979). Additionally, 5253 transcripts showed higher expression in S3. Pathway analysis showed 71 GO terms differentially enriched, with transmembrane receptor signaling pathway (GO:0007169) and phenylpropanoid metabolic/biosynthetic

processes (GO:0009698/GO:0009699) being the major biological processes differentially regulated. The lists of differentially expressed transcripts, as well as the enriched GO terms for all comparisons are available in Additional files 7 and 8, respectively.

Many genes involved in photosynthesis and phenylpropanoid related pathways were found to be differentially expressed between stages 2 and 3, and they were involved in distinct biological processes (Table 4). The list of individual genes involved in each biological process is available in Additional file 9. It has been shown that, during bract development in poinsettia, photosynthetic pigments are synthesized early and then replaced by different phenolic compounds [27, 36]. Thus, a significantly lower expression of genes related to photosynthesis, accompanied by a higher expression of flavonoid biosynthesis genes (phenylpropanoid pathway), was expected along with this transition.

Characterization of expression differences between red and white poinsettia varieties

For the characterization of the differences between Christmas Feelings and Christmas Feelings Pearl, three independent biological replicates were used for each of the varieties, and the comparison was performed for the three bract development stages. The pairwise comparison revealed 1204 transcripts with a lower expression in white bracts on the first stage, while only 130 were lower expressed on stage two and 673 on stage three (FDR ≤ 0.05). However, only 48 transcripts were equally lower expressed in white bracts for all stages (Fig. 2a). On the other hand, 1446 transcripts were higher expressed in white bracts on the first stage, whilst a lower number of higher expressed transcripts were detected on stages two and three (321 and 790, respectively). Nonetheless, 23 were commonly high expressed in white bracts in all stages (Fig. 2b).

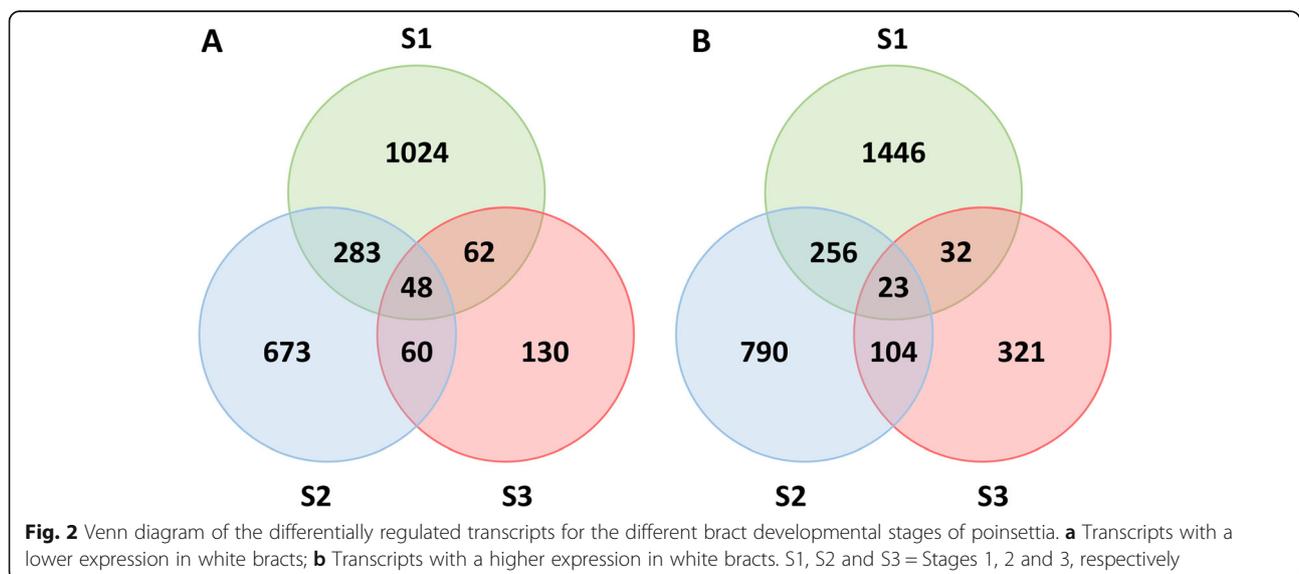
Pathway enrichment analysis was performed for the low- and high-expressed transcripts in white bracts for each of the developmental stages. Low expressed transcripts in the white bracts were associated with numerous biological processes. For stage one, 21 GO terms were differentially enriched, with major biological processes, such as response to temperature stimulus/heat (GO:0009266/GO:0009408) and flavonoid biosynthetic/metabolic process (GO:0009813/GO:0009812), among those. On the second stage, 11 GO terms were differentially enriched, with phosphorylation (GO:0016310) and protein phosphorylation (GO:0006468) among the major enriched biological processes pathways. As for the last stage, 10 GO terms were differentially enriched, with multidimensional cell growth (GO:0009825) and plant-type cell wall modification (GO:0009827) among the enriched biological processes.

Table 4 Differentially enriched photosynthesis- and phenylpropanoid-related pathways between stages 2 and 3 of poinsettia bract development

Down-regulated in stage 3			
GO term	Term description	Genes identified	FDR
GO:0015979	Photosynthesis	51	1.5E-10
GO:0006091	Generation of precursor metabolites and energy	42	0.0033
GO:0009657	Plastid organization	39	1.7E-07
GO:0019684	Photosynthesis, light reaction	34	3.5E-09
GO:0009658	Chloroplast organization	22	0.0087
GO:0009767	Photosynthetic electron transport chain	13	0.0054
GO:0045036	Protein targeting to chloroplast	10	0.0087
GO:0010027	Thylakoid membrane organization	10	0.0087
GO:0072598	Protein localization to chloroplast	10	0.0087
GO:0072596	Establishment of protein localization to chloroplast	10	0.0087
GO:0009668	Plastid membrane organization	10	0.0095
GO:0009773	Photosynthetic electron transport in photosystem I	8	0.0029
GO:0010207	Photosystem II assembly	7	0.0190
GO:0045038	Protein import into chloroplast thylakoid membrane	5	0.0100
Up-regulated in stage 3			
GO:0009698	Phenylpropanoid metabolic process	27	0.0490
GO:0009699	Phenylpropanoid biosynthetic process	25	0.0059

In the same way, various biological processes were linked with the higher expressed transcripts in the white bracts. For the first stage, a total of 99 GO terms were found to be differentially enriched, with photosynthesis (GO:0015979 - photosynthesis / GO:0019684 - photosynthesis, light reaction / GO:0009767 - photosynthetic electron transport chain) and abiotic stimulus (GO:0009416 - response to light stimulus / GO:0009314 - response to radiation / GO:0009409 - response to cold) among those enriched

pathways. As for the second stage, high expressed transcripts were involved in 62 differentially enriched GO terms. The main biological processes with a differential regulation were response to stimulus (GO:0050896), response to stress (GO:0006950), as well as phenylpropanoid biosynthetic/metabolic processes (GO:0009699/ GO:0009698). Lastly, 31 enriched GO terms were associated with the higher expressed transcripts in stage three. The main enriched biological processes were response to wounding (GO:0009611)



and jasmonic acid biosynthetic/metabolic processes (GO:0009695/GO:0009694). Moreover, several molecular functions related to transferase and glucosyl-transferase/glycosyltransferase activities (GO:0016757/GO:0008194/GO:0046527) were also enriched. The lists of differentially expressed transcripts, as well as the enriched GO terms for all comparisons are available in Additional files 10 and 11, respectively.

To further investigate possible differences in flavonoid biosynthesis genes, we analyzed the differentially expressed genes belonging to flavonoid metabolic process (GO:0009812) for each of the bract developmental stages between red and white poinsettia varieties. The main genes involved in the flavonoid biosynthesis and their difference in expression for each of the bract developmental stages are shown in Fig. 3. For the first stage of bract development, a

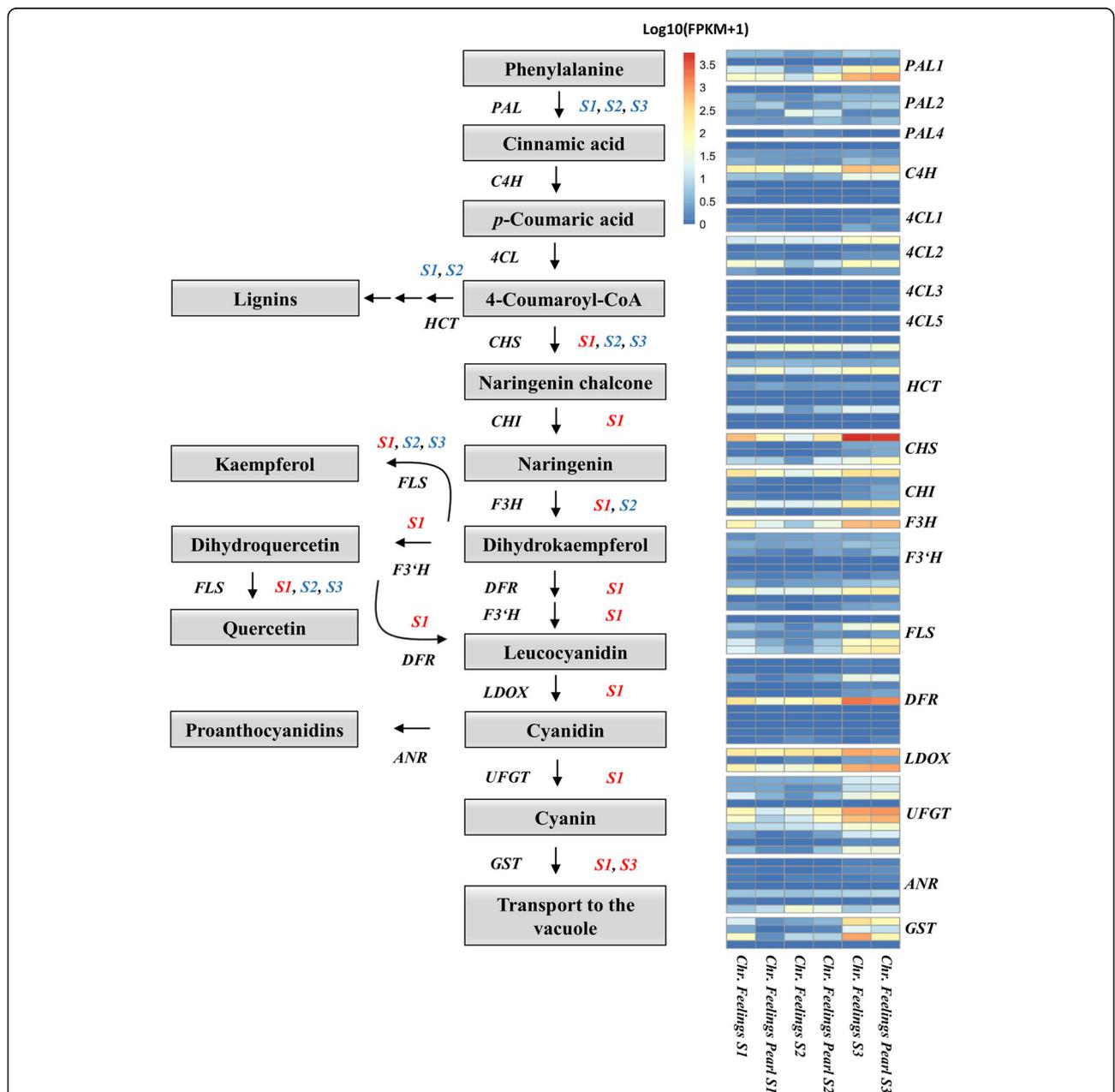


Fig. 3 Anthocyanin biosynthetic pathway and expression of related genes during bract development in poinsettia varieties. (left) Differentially expressed genes (FDR ≤ 0.05) in the three stages of bract development are depicted by S1, S2 and S3 (Stages 1, 2 and 3, respectively) symbols next to the genes. Stages colored in red indicate a higher expression of the respective gene in the red poinsettia variety. Stages colored in blue indicate a higher expression of the respective gene in the white poinsettia variety. (right) Heatmap of the genes involved in each process of the pathway. Gene expression is represented by Log10(FPKM+ 1). FPKM = Fragments per kilobase per million. For gene abbreviations refer to Table 2

total of 13 flavonoid-related genes showed differences in expression rates between red and white varieties, with 11 of them being lower expressed in the white variety (*CHS*, *CHI*, *F3H* (synonym: *FHT*), *F3'H*, *FLS1*, *DFR*, *LDOX*, *UGFT*, *MYB12*, *MYB113*, and *GSTF11*), while two of them showed a higher expression (*HCT* and *PAL2*). On the second stage, *PAL1*, *PAL2*, *HCT*, *CHS*, and *F3H* showed a higher expression in the white variety. For the last stage of bract development, five genes displayed differential expression between red and white varieties, with *GSTF11* being low expressed in the white variety, while *CHS*, *FLS*, *PAL2*, and *BEN* showed higher expression.

Two genes related to flavonoid biosynthesis showed antagonistic expression patterns along the bract development stages. *CHS* was lower expressed in white samples at the first stage, whereas in the second and third stages its expression was higher in white samples. As previously shown (Table 2), four transcripts were annotated as *CHS* in our bract transcriptome (here named *CHS1* to *CHS4*). *CHS1* was low expressed in the white variety in the first stage, but higher expressed in the second stage. In addition, *CHS2* was higher expressed in the white variety in the second and third stages. Similar results were identified for *FLS*, where five different transcripts were annotated as this gene in our transcriptome (here named *FLS1* to *FLS5*). *FLS1* and *FLS2* were lower expressed in white varieties on the first stage, while *FLS4* showed a higher expression in the last stage. Thus, the expression of some enzymes related to flavonoid biosynthesis might be driven by the complementary expression of multiple isoforms.

Validation of gene expression patterns by RT-qPCR validation

To further verify the expression profiles in the Illumina sequencing analyses, 10 transcripts were selected for RT-qPCR using the Christmas Feelings and Christmas Feelings Pearl varieties for each of the developmental stages used for RNA-Seq. The same biological triplicates used for RNA-Seq plus two extra independent biological samples were used for the RT-qPCR reactions. The selected genes are known to be part of the flavonoid and anthocyanin pathways in plants: *CHS*, *F3H*, *F3'H*, *DFR*, *ANR*, *LDOX*, *UGT79B10*, *UGT78D2*, *GSTF11*, and *GSTU17*. The normalized relative quantity (NRQ) obtained by RT-qPCR for each of the genes in the different time points and color bracts is shown in Fig. 4a. NRQ values were calculated relative to one of the biological replicates of the Christmas Feelings variety in stage 1 of bract development according to the Pfaffl method and equations [60]. In addition, the RNA-Seq expression for each of the genes is shown in Fig. 4b.

Most of the genes analyzed by RT-qPCR showed a similar expression trend to the RNA-Seq data. *ANR* was

the only analyzed gene that showed a completely different pattern of expression. The RT-qPCR primers were designed based on one of the transcripts annotated as an *A. thaliana ANR* homolog. However, several other transcripts have also been annotated as such (Table 2), with some of them showing distinct expression values among samples (data not shown), but none of them showing a differential expression on the RNA-Seq datasets. Moreover, other non-annotated transcripts might also have similarities to the designed primers and, therefore, might have been amplified in the RT-qPCR reaction. Nevertheless, these results indicate that the sequencing data produced in this study were accurate and reliable.

Discussion

Transcriptome assembly and annotation

Poinsettia is a widely popular ornamental plant, especially during the Christmas period, due to its red bract coloration. For the past years, a range of cultivars has been available, which exhibit differences mainly in height, growth habit, leaf size, and bract coloration. An understanding of the molecular mechanisms underlying bract development, particularly in color development and accumulation, will assist in the poinsettia breeding process to improve its ornamental value. However, scarce genetic information is available for the species. Complete genomes are only available for species from the same family, such as *Ricinus communis* [20], *Jatropha curcas* [66], *Manihot esculenta* [61] and *Hevea brasiliensis* [64], as well as some transcriptomes of *Euphorbia* species [9, 18, 32, 37, 62]. A recent transcriptome study has reported the assembly of 232,663 contigs arising from green leaf and red-turning bract of poinsettia [30], which is very similar to our transcriptome assembly (288,524 contigs). However, no functional annotation of the aforementioned transcriptome is available for comparison.

By applying the BUSCO pipeline, we confirmed that our transcriptome contains around 77% of the available ortholog groups at OrthoDB v9.1 [93]. Transcriptome studies with other plant species have shown a higher level of completeness (e.g. *Cinnamomum longepaniculatum* - 91% and *Noccaea caerulea* - 90% [13, 90]), while others are similar to the ones in our transcriptome (e.g. *Camellia nitidissima* - 76% [101]). Moreover, different levels of BUSCO completeness were observed when comparing different tissues of the same species [8], thus indicating that tissue-specific transcripts may account for different coverages compared to what is expected for the complete gene space. Nonetheless, when comparing our results to the leaf, stem and root transcriptomes of *E. pekinensis* assembled in this study, comparable levels of BUSCO completeness were observed, as well as the presence of tissue-specific ortholog groups.

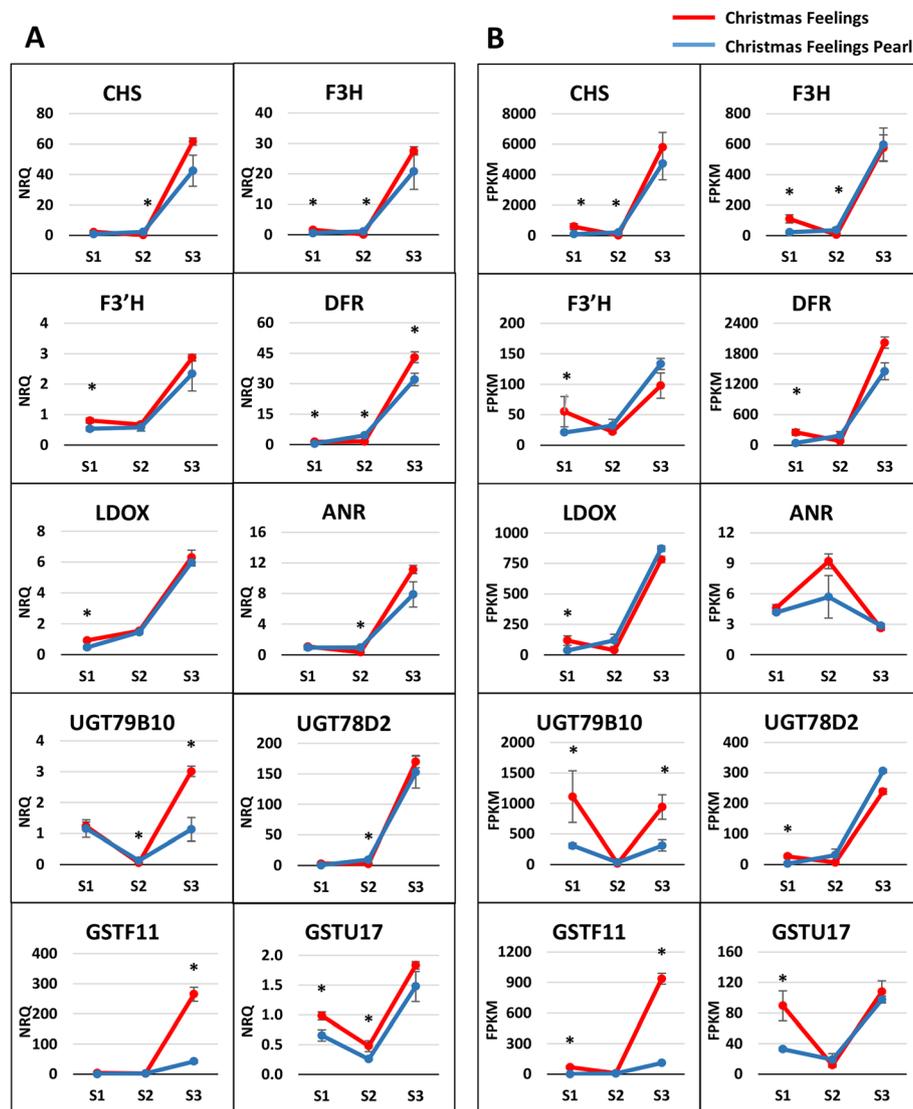


Fig. 4 Expression profiles of anthocyanin-related genes for three developmental stages of poinsettia bracts. **a** RT-qPCR expression profiles of 10 anthocyanin related genes for the varieties Christmas Feelings and Christmas Feelings Pearl in three stages of bract development. **b** RNA-Seq expression profiles of 10 anthocyanin related genes for the varieties Christmas Feelings and Christmas Feelings Pearl in three stages of bract development. S1, S2, S3 = Stages 1, 2 and 3, respectively. Vertical bars indicate standard errors. “*” symbol indicates significant differences for that specific stage for $p \leq 0.05$. FPKM = Fragments per kilobase per million. NRQ = Normalized relative quantity. For gene abbreviations refer to Table 2

In this study, we used a hybrid de novo assembly strategy (Illumina and PacBio platforms) to generate a transcriptome for poinsettia bracts, where 95,900 out of 288,524 contigs were confidently annotated against *A. thaliana* transcripts (TAIR10). These represent a set of 14,623 distinct *A. thaliana* homologous transcripts. The 192,624 contigs without annotation might represent family- or species-specific transcripts, but also short and incomplete transcripts; nonetheless, they need to be further analyzed in order to confirm their origin. Overall, these results will significantly enhance the available data for poinsettia in the public databases and will provide useful genetic information that could be exploited for breeding purposes.

Modulation of bract development

The flowering behavior of plants is regulated by distinct environmental aspects, with light playing a crucial role in several ways. Day-length, or photoperiod, regulates flowering time and allows sexual reproduction to happen at favorable times [73]. Plants are classified according to photoperiodic responses into long-day (LD), in which flowering occurs when the day becomes longer than some crucial length, and short-day (SD), in which flowering occurs when the day becomes shorter [33]. Photoperiod also plays an important role in regulating the biosynthesis of secondary metabolites in plants [34], with longer photoperiods generally promoting anthocyanin

biosynthesis [11, 49]. Nonetheless, some plants are able to activate the biosynthesis of anthocyanins in short photoperiod situations. Anthocyanin promotion has been observed in *A. thaliana* due to short photoperiod sensing by phytochrome A [67]. In *Begonia semperflorens*, short-day period, together with low temperatures, is crucial for anthocyanin biosynthesis and it is directly related to increased activities of the enzymes *PAL*, *CHI*, *DFR* and *UFGT* [95].

The flower formation in poinsettia, leading to bract formation and coloration, is induced under short day conditions [41], thus also indicating the role of photoperiodism in anthocyanin induction for the species. The bracts of poinsettia are leaves changing their photosynthetic function into pollinator attraction (i.e. by accumulating anthocyanins) upon flower induction to escort the relatively small and unimpressive reproductive structures [31, 57]. During the bract development process in poinsettia, especially between stages 2 and 3, several photosynthesis related pathways showed a down-regulation in the latest stage, followed by an up-regulation of phenylpropanoid related pathways (Table 4). Increased anthocyanin content levels were detected in the transition from partially to fully pigmented poinsettia bracts, which was accompanied by the reduction of photosynthetic pigments [7, 68]. Moreover, accumulation of chlorophyll was reduced when young poinsettia leaves started to accumulate anthocyanins under short day conditions, which was due to a decrease in the activity of enzymes related to chlorophyll synthesis [36]. In conclusion, the development of poinsettia bracts is marked by a decrease in photosynthesis and chlorophyll biosynthesis genes, followed by increased activity of genes related to flavonoid biosynthesis.

Regulation of flavonoid pathway between red and white poinsettia varieties during bract development

The anthocyanin biosynthetic pathway is a well-characterized and conserved network in plants, whose regulation is maintained through the expression of structural and regulatory biosynthetic genes [48]. The structural genes can be divided into early biosynthetic genes (EBGs), i.e. *CHS*, *CHI*, *F3H*, *F3'H*, *FLS*, and late biosynthetic genes (LBGs), i.e. *DFR*, *ANS/LDOX*, *UFGT*, *LAR*, *ANR* [22, 59]. EBGs are usually regulated by *R2R3-MYB* regulatory genes, whereas the activation of LBGs is mediated by a regulatory complex, called the MYB-bHLH-WD40 (MBW) complex, consisting of MYB, basic helix-loop-helix (bHLH) and WD40 repeat families [48, 59].

Our gene expression comparison between red and white varieties of poinsettia showed a differential regulation of the flavonoid biosynthesis pathway only at particular stages of bract development. Several structural genes showed a down-regulation on the white variety on the first analyzed stage. Interestingly, two *R2R3-MYB*

regulatory genes were also shown to be down-regulated in the white variety: *MYB12* and *MYB113*. *MYB11*, *MYB12*, and *MYB111* from *A. thaliana* share significant structural similarity and are involved in the regulation of the expression of EBGs [59, 70]. In *A. thaliana myb12*-ko mutant seedlings, *CHS* and *FLS* expressions showed a clear reduction, while the expression of *CHI*, *F3H*, *DFR*, and *F3'H* remained unchanged. In contrast, overexpression of *MYB12* in seedlings led to an increased expression of *CHS*, *CHI*, *F3H* and *FLS* [51]. MYB factors have also been demonstrated to positively regulate the expression of EBGs in other species [1, 21, 79, 89].

On the other hand, *R2R3-MYB* factors such as *PAP1*, *PAP2*, *MYB113*, *MYB114* are known to participate in the MBW complex and to regulate the expression of LBGs [10, 28]. In apple, the *MdMYB10* gene, a *MYB113* homologous, showed a positive expression correlation with anthocyanin accumulation, as well as with the expression of LBGs [23]. In *L. formosana*, the *LfMYB113* have been shown to directly activate the expression of two *DFR* homologous, thus promoting the anthocyanin synthesis in leaves [85]. Overexpression of bHLH and MYB-related transcription factor from snapdragon (*Antirrhinum majus*) in tomato fruits resulted in a higher expression of flavonoid-related genes (e.g. *F3'H*, *F3'5'H*, *ANS*, *UFGTs*), thus leading to a higher accumulation of anthocyanins [15].

Our results show an initial impairment of the flavonoid pathway early in the color accumulation process for the white poinsettia variety, but these differences were not observed in the subsequent stages of bract development. In the comparisons between red and white varieties for stages 2 and 3, most of the previously down-regulated genes related to flavonoid biosynthesis did not show any differential expression. In fact, some of them showed an up-regulation in the white variety for those stages; however, a few of these genes contain multiple annotated transcripts (e.g. *CHS* and *FLS*) with different expression patterns. In fact, *CHS* has been shown to play a major role in anthocyanin biosynthesis in different species, in which the appearance of white flowers or flower segments is driven by a lack of its expression [26, 53, 58, 71]. *CHS*, a well-characterized enzyme with a key role in the early steps of flavonoid biosynthesis, is known to be encoded by a multigene family in many plant species [81, 88]. In turnip, six *CHS* genes were identified, but only three of them were shown to be functional and to promote anthocyanin biosynthesis [100]. Three *CHS* genes have been characterized in *Citrus* and they have been shown to contribute differently and complementarily to the production of flavonoids [82]. Two out of four *CHS* identified in our bract transcriptome showed a differential expression between red and white varieties. However, this does not seem to affect the overall functionality of the flavonoid

pathway in the poinsettia varieties analyzed in our study, since the pigmentation of bracts is due to the accumulation of flavonoid compounds [69]. Taking all together, the initial impairment observed for the flavonoid pathway does not seem to be responsible for the lack of anthocyanin accumulation in white poinsettia bracts. This is confirmed by the constitutive expression of EBGs and LBGs in stages 2 and 3 of bract development.

The last step of the anthocyanin biosynthesis is characterized by the transfer of the glucosyl moiety from UDP-glucose to the 3-hydroxyl group of anthocyanidins by *UDP glucose: flavonoid 3-O-glucosyltransferase (UFGT)*, which results in the formation of stable colored pigments of anthocyanins 3-O-glucosides, as well as providing stability and water solubility in the plant [92, 99]. *UFGT* expression has been positively linked with anthocyanin accumulation in grapes and apples [39, 52]. In *A. thaliana*, *UGT78D2* (At5g17050) and *UGT75C1* (At4g14090) are the main genes suggested to be involved in the modification of the sugar moieties of anthocyanins, but with *UGT79B1* (At5g54060) having similar functions [42, 75, 91]. In our dataset, we identified a *UGT79B10* gene being up-regulated in the red variety at stage 3, which is highly similar to the *UGT79B1* gene and, therefore, might be also involved in the anthocyanin formation in poinsettia.

After biosynthesis, most conjugated flavonoids are transported and deposited primarily to the vacuole [45, 86], where vacuolar pH and the presence of co-pigments determine anthocyanin-mediated coloration [98]. Three distinct mechanisms for flavonoid transport in plant cells have been proposed: vesicle trafficking, membrane-mediated transport, and Glutathione S-transferase (GST) mediated transport [98]. GST genes play an important role in anthocyanin transportation, since GST mutants show phenotypes with a visible lack of pigmentation, such as *bz2* (*Bronze-2*) from maize, *an9* (*Anthocyanin 9*) from petunia, *tt19* (*Transparent Testa 19*) from Arabidopsis and *fl3* (*Flavonoid3*) from carnation [4, 38, 44, 50]. Moreover, there is a high conservation of GSTs involved in flavonoid accumulation [97] and, therefore, they are able to complement each other's expression.

In our differential expression analysis, a *GSTF11* Arabidopsis homolog gene showed a higher expression in the red variety for the last stage of bract development for both RNA-Seq and RT-qPCR analyses (Fig. 4). Although *GSTF12* is shown to be involved in anthocyanin transport [38], they share a high similarity. In fact, the poinsettia putative *GST* gene shares 58 and 55% amino acid identity with Arabidopsis *GSTF11* and *GSTF12*, respectively, which is higher than between Arabidopsis *TT19* and petunia *AN9* (50% amino acid identity) [38]. Due to its homology to known anthocyanin-related GSTs, the putative poinsettia *GST* is a promising candidate for white coloration in poinsettia.

Conclusions

In this study, we provide a comprehensive hybrid transcriptome from poinsettia bracts. In addition, we provide for the first time a profiling of gene expression during the process of bract development of red and white poinsettia varieties. Our differential expression analysis revealed that the majority of the anthocyanin-related genes are equally expressed in red and white varieties. Nonetheless, *UGT79B10* and *GSTF11* showed a lower expression in the last stage of bract development for the white variety, which are involved in glucosylation and transport of anthocyanins. The role of the putative *UGT79B10* and *GST* in the differences in anthocyanin accumulation in red and white poinsettias is still unknown. Functional studies are needed in order to clarify their possible role in the transition from red to white bracts. Nonetheless, these genes, and genes regulating their expression, are potential candidates for further studies.

Our transcriptome analysis provides a valuable foundation for further studies on the species, such as plant breeding and genetics, and highlights crucial information on the molecular mechanism of color formation in poinsettia. It should promote further investigations into the detailed regulatory pathways regulating flavonoid biosynthesis and contribute to a better understanding of the *white paradox* in the species.

Methods

Plant material and growth conditions

The red bracted poinsettia varieties Vintage and Christmas Feelings, as well as the white bracted variety Christmas Feelings Pearl were used in the present study. The white variety was obtained through radiation mutagenesis of the red variety, followed by shoot development and trait selection at the company Selecta One (Stuttgart, Germany). Therefore, red- and white-bracted poinsettias from the same variety are referred to as 'pairs', due to their highly similar genetic background. The varieties' names, bract color, number of biological replicates and other aspects are shown in Table 5. Bract samples were harvested in three color developmental stages: i) Stage 1 (S1) – defined as the transition of green colored leaves to red/white colored bracts, with the majority of the bracts still bearing a greenish coloration; ii) Stage 2 (S2) – defined as the presence of both green and red/white colors in the bracts, with a major part of the bracts bearing red/white coloration; and iii) Stage 3 (S3) – defined as a fully developed red/white coloration, with no major green coloration visible on the bracts. For a visual representation of the stages, please refer to Fig. 5.

Rooted cuttings from all varieties were obtained from Selecta One company (<https://www.selecta-one.com/>) and grown in a greenhouse, at the Institute for Plant

Table 5 Pairs of red and white poinsettia varieties used in the present study

Type of analysis	Variety name	Bract coloration	Color stage	# of biological replicates
Illumina RNA-Seq <i>single-end</i>	Christmas Feelings	Red	S1	3
	Christmas Feelings Pearl	White	S1	3
	Christmas Feelings	Red	S2	3
	Christmas Feelings Pearl	White	S2	3
	Christmas Feelings	Red	S3	3
	Christmas Feelings Pearl	White	S3	3
Illumina RNA-Seq <i>paired-end</i>	Christmas Feelings	Red	S3	1
	Christmas Feelings Pearl	White	S3	1
PacBio RNA-Seq	Vintage	Red	S3	1
RT-qPCR	Christmas Feelings	Red	S1	5
	Christmas Feelings Pearl	White	S1	5
	Christmas Feelings	Red	S2	5
	Christmas Feelings Pearl	White	S2	5
	Christmas Feelings	Red	S3	5
	Christmas Feelings Pearl	White	S3	5

Genetics from the Leibniz Universität Hannover (Hannover, Germany), under short-day conditions to induce flower formation and to stimulate the development of colored bracts. The plants were grown in 5 L pots containing Einheitserde P substrate (Hermann Meyer KG, Germany), with an average temperature of 22 °C and 9 h of daylight (15 h of darkness). Bract samples were harvested, immediately frozen in liquid nitrogen and stored at – 80 °C for subsequent analysis.

Tissue sampling, RNA isolation, and quantification

Bract samples from all varieties used for RNA-Seq were sent on dry ice to vertis Biotechnologie AG (Freising, Germany) for processing. Total RNA was isolated from approximately 100 mg of bract tissue using the mirPremier™ miRNA isolation kit (Sigma-Aldrich, St. Louis, USA) according to the manufacturer's instructions. Total RNA samples were analyzed for integrity by capillary electrophoresis using Shimadzu MultiNA microchip electrophoresis MCE-202 MultiNA Microchip Electrophoresis System (Shimadzu Corp., Kyoto, Japan).

For RT-qPCR analysis, total RNA was isolated from approximately 100 mg of bract tissue using the mirPremier™ miRNA isolation kit (Sigma-Aldrich) at the Institute for Plant Genetics from the Leibniz Universität Hannover. The total RNA concentration was analyzed using NanoDrop™ 2000 (Thermo Fisher Scientific, Wilmington, USA) and agarose gel electrophoresis.

PacBio sequencing and data processing

A full-length cDNA library from the Vintage variety was prepared at vertis Biotechnologie AG. Briefly, Poly(A) +

RNA was isolated from the total RNA sample and the 5'CAP structure was removed using CAP-Clip™ Acid Pyrophosphatase (Cellsript, Wisconsin, USA). Afterward, an RNA adapter was ligated to the 5'-monophosphate of the RNA. First strand cDNA was synthesized using an oligo (dT)-linker primer and M-MLV [H–] Reverse Transcriptase (Promega, Wisconsin, USA). The library sequencing was performed at the Vienna BioCenter Core Facilities GmbH (Vienna, Austria) using the PacBio Sequel System based on the Single Molecule, Real-Time (SMRT) Sequencing technology.

The Isoform Sequencing (Iso-Seq) Analysis v3.0 pipeline (<https://github.com/ben-lerch/IsoSeq-3.0>) was used to analyze the PacBio dataset. The pipeline was performed in three stages: i) CCS, where circular consensus sequences (CCS) were built from subreads; ii) Classify, where CCSs were classified as full-length non-chimeric (FLNC) reads and non-full length (NFL) reads; and iii) Cluster, where the sequences were clustered in high-quality consensus sequences (contigs).

Illumina sequencing and data processing

Two different sequencing strategies were used for the Illumina sequencing. In the first one, 1x75bp *single-end* 3' cDNA libraries were constructed for the varieties Christmas Feelings and Christmas Feelings Pearl for the different bract developmental stages. Poly(A) + RNA was isolated from the total RNA samples and the first-strand cDNA was synthesized using an oligo (dT)-adapter primer and M-MLV reverse transcriptase. After fragmentation, the first-strand cDNA was purified, the 5' Illumina TruSeq sequencing adapter

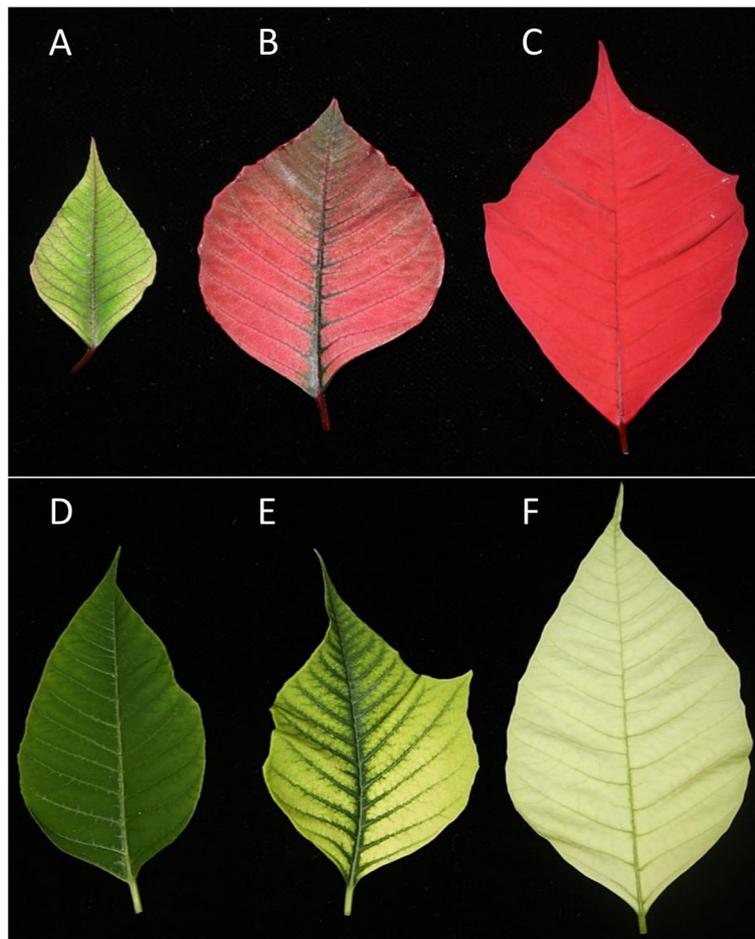


Fig. 5 Bracts of red and white poinsettia varieties for three color developmental stages. **a-c** Bracts from the Christmas Feelings variety for stages 1, 2 and 3, respectively; **d-f** Bracts from the Christmas Feelings Pearl variety for stages 1, 2 and 3, respectively

was ligated to the 3' end of the antisense cDNA and, finally, amplified by PCR.

For the second strategy, 2x150bp *paired-end* cDNA libraries were constructed for the varieties Christmas Feelings and Christmas Feelings Pearl for the third stage of bract development (S3). Ribosomal RNA molecules were depleted using the Ribo-Zero rRNA Removal Kit for plants (Illumina, San Diego, USA). Second, the first-strand cDNA was synthesized using an N6 randomized primer. After fragmentation, the Illumina TruSeq sequencing adapters were ligated in a strand-specific manner to the 5' and 3' ends of the cDNA fragments and the cDNA was finally amplified by PCR. Both *paired-end* and *single-end* libraries were sequenced at vertis Biotechnologie AG using an Illumina NextSeq500 system.

Reads representing ribosomal RNA gene fragments (rRNAs) were removed from the datasets using the sortmerna tool v2.1 [40] with all included databases: SILVA and Rfam [35, 63]. Reads were trimmed and filtered using Trimmomatic v0.36 [14] with the parameters

adapted to both sequencing strategies: 2x150bp *paired-end*: TRAILING:20 AVGQUAL:20 SLIDINGWINDOW:5:20 MINLEN:75; 1x75bp *single-end*: TRAILING:20 AVGQUAL:20 SLIDINGWINDOW:5:20 MINLEN:50.

Transcriptome assembly, annotation, and completeness of the transcriptome

The poinsettia bract transcriptome was assembled using the high-quality PacBio consensus sequences and the 150 bp *paired-end* processed Illumina reads from Christmas Feelings and Christmas Feelings Pearl varieties. The assembly was performed with Trinity v2.7.0 [29] using the long-reads assembly option. The transcriptome was annotated by sequence similarity against the *Arabidopsis thaliana* genome (TAIR10 protein representative gene model) [12] and the SwissProt databases [77] using BLASTX v2.8.0 (*E-value* < 1E-20) [16]. GO terms were retrieved, for the final poinsettia bract transcriptome, from the best hits obtained from BLASTX against the TAIR10 database using the online tool agriGO v2.0 [74].

Additionally, the BUSCO pipeline v1.2 [83] with its plant set (embryophyta_odb9) was used to assess the completeness of the poinsettia bract transcriptome.

For understanding the sequence and quantitative differences between tissue-specific transcripts in *Euphorbia* species, short *paired-end* Illumina RNA sequences from *Euphorbia pekinensis* root, stem, and leaf tissues were retrieved from the NCBI Sequence Read Archive (SRA) Sequence Database (accession number SRP097008) [18]. Ribosomal RNAs were removed from the datasets using the sortmerna tool v2.1 [40], followed by low-quality reads (average quality score below 20) trimming using Trimmomatic v0.36 [14] with the parameters TRAILING:20 AVGQUAL:20 SLIDINGWINDOW:5:20 MINLEN:75. De novo transcriptomes were assembled for each of the tissues using Trinity v2.7.0 [29]. Annotation and retrieval of GO terms for each of the tissues' transcriptomes were done in a similar way as for the poinsettia transcriptome.

Differential gene expression and pathway enrichment analysis

Illumina processed reads from the different red and white poinsettia samples were used for the differential gene expression (DGE) analysis. Transcript abundance quantification was performed with the RSEM tool [46] and bowtie2 was selected as the alignment method. Low expressed transcripts (Counts Per Million (CPM) ≤ 0.5 in at least 2 biological replicates) were removed from the dataset. Normalizations and pair-wise comparisons were performed with edgeR [65]. The thresholds for a differentially expressed gene (DEG) were set as: i) False Discovery Rate (FDR) ≤ 0.05 ; ii) $\log_2\text{FC} \geq 1$ or ≤ -1 ; and iii) Fragments Per Kilobase of transcript per Million mapped reads (FPKM) ≥ 1.0 for three biological replicates in at least one of the compared stages.

The differentially expressed genes for each of the comparisons were subjected to Single Enrichment Analysis (SEA) using the online tool agriGO v2.0, with the following parameters: 1) Selected species: *Arabidopsis thaliana*; 2) Reference: TAIR genome locus (TAIR10_2017); 3) Statistical test method: Hypergeometric; 4) Multi-test adjustment method: Hochberg (FDR); 5) Significance level of 0.05; 6) Minimum number of 5 mapping entries; and 7) Gene ontology type: Complete GO.

Quantitative PCR

cDNA synthesis was performed using the FastGene Scriptase Basic cDNA Kit (Nippon Genetics Europe GmbH, Düren, Germany) according to the manufacturer's recommendations. A total of five independent biological replicates were used for each of the varieties and stages analyzed (Table 5). The RT-qPCRs were performed using the qPCRBIO SyGreen Mix Lo-ROX kit

(Nippon Genetics Europe GmbH) according to the manufacturer's recommendations. Briefly, reactions were carried out in technical triplicates in a volume of 10 μL containing 5 μL of qPCRBIO SyGreen Mix Lo-ROX, 10 μmol of gene-specific forward and reverse primers, and 4 μL of 1:50 cDNA dilution. RT-qPCRs were performed using a StepOne™ Real-Time PCR System (Applied Biosystems, Singapore, Singapore). The normalized relative quantity (NRQ) was calculated according to the Pfaffl equations [60]. Two reference genes (Translation elongation factor 1 beta – *EF1B*; and Translation elongation factor 1 alpha – *EF1A* [94]) were used to normalize the expression data. The list of genes and primer sequence-design for the RT-qPCR reactions are available in Additional file 12. Statistical analysis was performed using the Relative Expression Software Tool (REST) v2.0.13 [60].

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6247-3>.

Additional file 1. Histogram of average coverage of *paired-end* Illumina reads from Christmas Feelings (A) and Christmas Feelings Pearl (B) varieties mapped to the 30,768 PacBio contigs.

Additional file 2. Annotation results of the poinsettia bract transcriptome against the TAIR10 database.

Additional file 3. Distribution of percent length coverage for the top matching database entries from the poinsettia bract transcripts.

Additional file 4. BUSCO results for the *E. pulcherrima* and *E. pekinensis* transcriptomes.

Additional file 5. Summary of the processed libraries for the Christmas Feelings and Christmas Feelings Pearl poinsettia varieties in three bract developmental stages.

Additional file 6. Pearson correlation for the biological replicates of Christmas Feelings and Christmas Feelings Pearl *paired-end* Illumina datasets for the three stages of bract development in poinsettia.

Additional file 7. Differentially expressed transcripts for the comparisons between bract developmental stages in poinsettia.

Additional file 8. List of enriched GO terms for the comparisons between bract developmental stages in poinsettia.

Additional file 9. List of transcripts present in the differentially enriched pathways related to photosynthesis and phenylpropanoid metabolic processes.

Additional file 10. Differentially expressed transcripts for the comparisons between Christmas Feelings (red) and Christmas Feelings Pearl (white) poinsettia varieties for the different bract developmental stages.

Additional file 11. List of enriched GO terms for the comparisons between Christmas Feelings (red) and Christmas Feelings Pearl (white) poinsettia varieties for the different bract developmental stages.

Additional file 12. Primer sequences for each of the genes analyzed by RT-qPCR in poinsettia bracts.

Abbreviations

CCS: Circular consensus sequences; cDNA: Complementary DNA; CPM: Counts Per Million; DEG: Differentially expressed gene; DGE: Differential gene expression; DNA: Deoxyribonucleic acid; EBG: Early biosynthetic gene; FDR: False discovery rate; FLNC: Full-length non-chimeric reads; FPKM: Fragments Per Kilobase of transcript per Million mapped reads; GC

content: Guanine-cytosine content; GO: Gene ontology; GST: Glutathione S-transferase; LBG: Late biosynthetic gene; LD: Long-day; MBW complex: MYB-bHLH-WD40 complex; NCBI: National Center for Biotechnology Information; NFL: Non-full length reads; NRQ: Normalized relative quantity; PCR: Polymerase chain reaction; PPR: Pentatricopeptide repeat; RNA: Ribonucleic acid; RNA-Seq: RNA sequencing; ROS: Reactive oxygen species; rRNA: Ribosomal RNA; RT-qPCR: Quantitative reverse transcription PCR; SD: Short-day; SMRT: Single Molecule, Real-Time Sequencing technology; SRA: Sequence Read Archive

Acknowledgments

We would like to thank Christopher Schlosser for critical reading of the manuscript and Guido von Tubeuf for providing the poinsettia varieties.

Authors' contributions

HH, RB, TR, and TD conceived the research. W, CL, HH, RB, TR, and TD designed the experiments. W and CL conducted the bioinformatics analyses. W performed the RT-qPCRs. W and CL wrote the manuscript. HH, RB, TR, and TD revised the draft of the manuscript. All authors read and approved the final manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 675657 Flower Power. This funding body had no role in the design of the study, collection, analysis, or interpretation of data, or in writing the manuscript. The publication of this article was funded by the Open Access fund of Leibniz Universität Hannover.

Availability of data and materials

The full sequencing dataset (Illumina and PacBio) is available through the Sequence Read Archive (SRA) at NCBI under BioProject number PRJNA532349.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Plant Genetics, Leibniz Universität Hannover, 30419 Hannover, Germany. ²Klemm + Sohn GmbH & Co., 70379 Stuttgart, KG, Germany. ³Department of Microbiology and Ecosystem Science, University of Vienna, 1090 Vienna, Austria. ⁴Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, 1060 Vienna, Austria.

Received: 1 July 2019 Accepted: 30 October 2019

Published online: 27 November 2019

References

- Adato A, Mandel T, Mintz-Oron S, Venger I, Levy D, Yativ M, et al. Fruit-surface flavonoid accumulation in tomato is controlled by a SIMYB12-regulated transcriptional network. *PLoS Genet.* 2009;5:e1000777. <https://doi.org/10.1371/journal.pgen.1000777>.
- Ahmed NU, Park J-I, Jung H-J, Yang T-J, Hur Y, Nou I-S. Characterization of dihydroflavonol 4-reductase (DFR) genes and their association with cold and freezing stress in *Brassica rapa*. *Gene.* 2014;550:46–55. <https://doi.org/10.1016/j.gene.2014.08.013>.
- Ahn JH, Kim J-S, Kim S, Soh HY, Shin H, Jang H, et al. De novo Transcriptome analysis to identify anthocyanin biosynthesis genes responsible for tissue-specific pigmentation in *Zoysiagrass* (*Zoysia japonica* Steud.). *PLoS One.* 2015;10:e0124497. <https://doi.org/10.1371/journal.pone.0124497>.
- Alfenito MR. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-Transferases. *Plant Cell Online.* 1998;10:1135–50. <https://doi.org/10.1105/tpc.10.7.1135>.
- An D, Cao HX, Li C, Humbeck K, Wang W. Isoform sequencing and state-of-art applications for Unravelling complexity of plant Transcriptomes. *Genes.* 2018;9(1):43. <https://doi.org/10.3390/genes9010043>.
- Ardui S, Ameur A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res.* 2018;46:2159–68. <https://doi.org/10.1093/nar/gky066>.
- Arreola JA, González AMC, García EA, Pineda JP, Aguilar LAV, León MTC. Effect of calcium, boron and molybdenum on plant growth and bract pigmentation in poinsettia. *Rev Fitotec Mex.* 2008;31:165–72.
- Babineau M, Mahmood K, Mathiassen SK, Kudsk P, Kristensen M. De novo transcriptome assembly analysis of weed *Apera spica-venti* from seven tissues and growth stages. *BMC Genomics.* 2017;18:128. <https://doi.org/10.1186/s12864-017-3538-4>.
- Barrero RA, Chapman B, Yang Y, Moolhuijzen P, Keeble-Gagnère G, Zhang N, et al. De novo assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes. *BMC Genomics.* 2011;12:600. <https://doi.org/10.1186/1471-2164-12-600>.
- Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 2004;39:366–80. <https://doi.org/10.1111/j.1365-3113.2004.02138.x>.
- Beckwith AG, Zhang Y, Seeram NP, Cameron AC, Nair MG. Relationship of light quantity and anthocyanin production in *Pennisetum setaceum* Cvs. Rubrum and red riding hood. *J Agric Food Chem.* 2004;52:456–61. <https://doi.org/10.1021/jf034821>.
- Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The *Arabidopsis* information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* (New York, N.Y.: 2000). 2015;53:474–85. <https://doi.org/10.1002/dvg.22877>.
- Blande D, Halimaa P, Tervahauta AI, Aarts MGM, Kärenlampi SO. De novo transcriptome assemblies of four accessions of the metal hyperaccumulator plant *Noccaea caerulea*. *Sci Data.* 2017;4:160131. <https://doi.org/10.1038/sdata.2016.131>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (Oxford, England). 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
- Butelli E, Titta L, Giorgio M, Mock H-P, Matros A, Peterek S, et al. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol.* 2008;26:1301–8. <https://doi.org/10.1038/nbt.1506>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>.
- Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, et al. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol Ecol Resour.* 2012;12:834–45. <https://doi.org/10.1111/j.1755-0998.2012.03148.x>.
- Cao X, Zhang F, Yuan B, Meng L, Yang X, Jiang J. De novo transcriptome sequencing and analysis of *Euphorbia pekinensis* Rupr. and identification of genes involved in diterpenoid biosynthesis. *Plant Gene.* 2017;12:33–42. <https://doi.org/10.1016/j.plgene.2017.07.001>.
- Chalker-Scott L. Environmental significance of Anthocyanins in plant stress responses. *Photchem Photobiol.* 1999;70:1. <https://doi.org/10.1111/j.1751-1097.1999.tb01944.x>.
- Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiui D, et al. Draft genome sequence of the oilseed species *Ricinus communis*. *Nat Biotechnol.* 2010;28:951–6. <https://doi.org/10.1038/nbt.1674>.
- Czemmel S, Stracke R, Weisshaar B, Cordon N, Harris NN, Walker AR, et al. The grapevine R2R3-MYB transcription factor VvMYB1 regulates flavonol synthesis in developing grape berries. *Plant Physiol.* 2009;151:1513–30. <https://doi.org/10.1104/pp.109.142059>.
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 2010;15:573–81. <https://doi.org/10.1016/j.tplants.2010.06.005>.
- Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC. Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 2007;49:414–27. <https://doi.org/10.1111/j.1365-3113.2006.02964.x>.
- Fang Z-Z, Zhou D-R, Ye X-F, Jiang C-C, Pan S-L. Identification of candidate anthocyanin-related genes by Transcriptomic analysis of 'Furongli' plum

- (*Prunus salicina* Lindl.) during fruit ripening using RNA-Seq. *Front Plant Sci.* 2016;7:1338. <https://doi.org/10.3389/fpls.2016.01338>.
25. Fu S, Ma Y, Yao H, Xu Z, Chen S, Song J, et al. IDP-denovo: de novo transcriptome assembly and isoform annotation by hybrid sequencing. *Bioinformatics.* 2018;34:2168–76. <https://doi.org/10.1093/bioinformatics/bty098>.
 26. Fukada-Tanaka S, Hoshino A, Hisatomi Y, Habu Y, Hasebe M, Iida S. Identification of new chalcone synthase genes for flower pigmentation in the Japanese and common morning glories. *Plant Cell Physiol.* 1997;38:754–8.
 27. González AMC, García EA, Pineda JP, Arreola JA, Aguilar LAV, León MTC. Effect of calcium, boron and molybdenum on plant growth and bract pigmentation in poinsettia. *Rev Fitotec Mex.* 2008;31(2):165–72.
 28. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* 2008;53:814–27. <https://doi.org/10.1111/j.1365-3113.2007.03373.x>.
 29. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52. <https://doi.org/10.1038/nbt.1883>.
 30. Gu Z, Chen H, Yang R, Ran M. Identification of DFR as a promoter of anthocyanin accumulation in poinsettia (*Euphorbia pulcherrima*, Willd. ex Klotsch) bracts under short-day conditions. *Sci Hortic.* 2018;236:158–65. <https://doi.org/10.1016/j.scienta.2018.03.032>.
 31. Horn JW, van Ee BW, Morawetz JJ, Riina R, Steinmann VW, Berry PE, et al. Phylogenetics and the evolution of major structural characters in the giant genus *Euphorbia* L. (*Euphorbiaceae*). *Mol Phylogenet Evol.* 2012;63:305–26. <https://doi.org/10.1016/j.ympev.2011.12.022>.
 32. Horvath DP, Chao WS, Suttle JC, Thimmapuram J, Anderson JV. Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (*Euphorbia esula* L.). *BMC Genomics.* 2008;9:–536. <https://doi.org/10.1186/1471-2164-9-536>.
 33. Imaizumi T, Kay SA. Photoperiodic control of flowering: not only by coincidence. *Trends Plant Sci.* 2006;11:550–8. <https://doi.org/10.1016/j.tplants.2006.09.004>.
 34. Jaakola L, Hohtola A. Effect of latitude on flavonoid biosynthesis in plants. *Plant Cell Environ.* 2010;33:1239–47. <https://doi.org/10.1111/j.1365-3040.2010.02154.x>.
 35. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 2018;46:D335–42. <https://doi.org/10.1093/nar/gkx1038>.
 36. Kannangara CG, Hansson M. Arrest of chlorophyll accumulation prior to anthocyanin formation in *Euphorbia pulcherrima*. *Plant Physiol Biochem.* 1998;36:843–8. [https://doi.org/10.1016/S0981-9428\(99\)80001-1](https://doi.org/10.1016/S0981-9428(99)80001-1).
 37. Kitajima S, Miura K, Aoki W, Yamato KT, Taira T, Murakami R, et al. Transcriptome and proteome analyses provide insight into laticifer's defense of *Euphorbia tirucalli* against pests. *Plant Physiol Biochem.* 2016;108:434–46. <https://doi.org/10.1016/j.plaphy.2016.08.008>.
 38. Kitamura S, Shikazono N, Tanaka A. TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *Plant J.* 2004;37:104–14. <https://doi.org/10.1046/j.1365-3113.2003.01943.x>.
 39. Kobayashi S, Ishimaru M, Ding CK, Yakushiji H, Goto N. Comparison of UDP-glucose:flavonoid 3-O-glucosyltransferase (UGFT) gene sequences between white grapes (*Vitis vinifera*) and their sports with red skin. *Plant Sci.* 2001;160:543–50. [https://doi.org/10.1016/S0168-9452\(00\)00425-8](https://doi.org/10.1016/S0168-9452(00)00425-8).
 40. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England).* 2012;28:3211–7. <https://doi.org/10.1093/bioinformatics/bts611>.
 41. Kristoffersen T. Influence of day length and temperature on the growth and development in poinsettia (*Euphorbia pulcherrima* Willd.). *Acta Hort.* 1969;14:73–89. <https://doi.org/10.17660/ActaHortic.1969.14.7>.
 42. Kubo H, Nawa N, Lupsea SA. Anthocyaninless1 gene of *Arabidopsis thaliana* encodes a UDP-glucose:flavonoid-3-O-glucosyltransferase. *J Plant Res.* 2007;120:445–9. <https://doi.org/10.1007/s10265-006-0067-7>.
 43. Landi M, Tattini M, Gould KS. Multiple functional roles of anthocyanins in plant-environment interactions. *Environ Exp Bot.* 2015;119:4–17. <https://doi.org/10.1016/j.envexpbot.2015.05.012>.
 44. Larsen ES, Alfenito MR, Briggs WR, Walbot V. A carnation anthocyanin mutant is complemented by the glutathione S-transferases encoded by maize Bz2 and petunia An9. *Plant Cell Rep.* 2003;21:900–4. <https://doi.org/10.1007/s00299-002-0545-x>.
 45. Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, et al. Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol.* 2006;57:405–30. <https://doi.org/10.1146/annurev.arplant.57.032905.105252>.
 46. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323. <https://doi.org/10.1186/1471-2105-12-323>.
 47. Li J, He Y-J, Zhou L, Liu Y, Jiang M, Ren L, et al. Transcriptome profiling of genes related to light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) before purple color becomes evident. *BMC Genomics.* 2018;19:201. <https://doi.org/10.1186/s12864-018-4587-z>.
 48. Liu Y, Tikunov Y, Schouten RE, Marcellis LFM, Visser RGF, Bovy A. Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: A review. *Front Chem.* 2018;6:52. <https://doi.org/10.3389/fchem.2018.00052>.
 49. Lu Y, Zhang M, Meng X, Wan H, Zhang J, Tian J, et al. Photoperiod and shading regulate coloration and anthocyanin accumulation in the leaves of malus crabapples. *Plant Cell Tissue Organ Cult.* 2015;121:619–32. <https://doi.org/10.1007/s11240-015-0733-3>.
 50. Marrs KA, Alfenito MR, Lloyd AM, Walbot V. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature.* 1995;375:397. <https://doi.org/10.1038/375397a0>.
 51. Mehtens F, Kranz H, Bednarek P, Weisshaar B. The *Arabidopsis* transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 2005;138:1083–96. <https://doi.org/10.1104/pp.104.058032>.
 52. Meng R, Qu D, Liu Y, Gao Z, Yang H, Shi X, et al. Anthocyanin accumulation and related gene family expression in the skin of dark-grown red and non-red apples (*Malus domestica* Borkh.) in response to sunlight. *Sci Hortic.* 2015;189:66–73. <https://doi.org/10.1016/j.scienta.2015.03.046>.
 53. Morita Y, Saito R, Ban Y, Tanikawa N, Kuchitsu K, Ando T, et al. Tandemly arranged chalcone synthase A genes contribute to the spatially regulated expression of siRNA and the natural bicolor floral phenotype in *Petunia hybrida*. *Plant J.* 2012;70:739–49. <https://doi.org/10.1111/j.1365-3113.2012.04908.x>.
 54. Moustaka J, Panteris E, Adamakis I-DS, Tanou G, Giannakoula A, Eleftheriou EP, et al. High anthocyanin accumulation in poinsettia leaves is accompanied by thylakoid membrane unstacking, acting as a photoprotective mechanism, to prevent ROS formation. *Environ Exp Bot.* 2018. <https://doi.org/10.1016/j.envexpbot.2018.01.006>.
 55. Nielsen M-D, Farestveit B, Andersen AS. Adventitious shoot development from decapitated plants of periclinal chimeric poinsettia plants (*Euphorbia pulcherrima* Willd ex Klotsch). *Eur J Hortic Sci.* 2003;68:161–8.
 56. Ning C, Cheng X, Luo P, Liang F, Wang Z, Yu G, et al. Hybrid sequencing and map finding (HySeMaFi): optional strategies for extensively deciphering gene splicing and expression in organisms without reference genome. *Sci Rep.* 2017;7:43793. <https://doi.org/10.1038/srep43793>.
 57. Nitarska D, Stefanini C, Haselmair-Gosch C, Miosic S, Walliser B, Mikulic-Petkovsek M, et al. The rare orange-red colored *Euphorbia pulcherrima* cultivar 'Harvest Orange' shows a nonsense mutation in a flavonoid 3'-hydroxylase allele expressed in the bracts. *BMC Plant Biol.* 2018;18:216. <https://doi.org/10.1186/s12870-018-1424-0>.
 58. Ohno S, Hosokawa M, Kojima M, Kitamura Y, Hoshino A, Tatsuzawa F, et al. Simultaneous post-transcriptional gene silencing of two different chalcone synthase genes resulting in pure white flowers in the octoploid dahlia. *Planta.* 2011;234:945–58. <https://doi.org/10.1007/s00425-011-1456-2>.
 59. Petroni K, Tonelli C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* 2011;181:219–29. <https://doi.org/10.1016/j.plantsci.2011.05.009>.
 60. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001;29:45e–45. <https://doi.org/10.1093/nar/29.9.e45>.
 61. Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, et al. The cassava genome: current Progress, future directions. *Trop Plant Biol.* 2012;5:88–94. <https://doi.org/10.1007/s12042-011-9088-z>.
 62. Qiao W, Li C, Mosongo I, Liang Q, Liu M, Wang X. Comparative Transcriptome Analysis Identifies Putative Genes Involved in Steroid Biosynthesis in *Euphorbia tirucalli*. *Genes.* 2018;9. <https://doi.org/10.3390/genes910038>.
 63. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-

- based tools. *Nucleic Acids Res.* 2013;41:D590–6. <https://doi.org/10.1093/nar/gks1219>.
64. Rahman AYA, Usharraj AO, Misra BB, Thottathil GP, Jayasekaran K, Feng Y, et al. Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics.* 2013;14:75. <https://doi.org/10.1186/1471-2164-14-75>.
 65. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England).* 2010;26:139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
 66. Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, et al. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res.* 2011;18:65–76. <https://doi.org/10.1093/dnares/dsq030>.
 67. Seaton DD, Toledo-Ortiz G, Ganpudi A, Kubota A, Imaizumi T, Halliday KJ. Dawn and photoperiod sensing by phytochrome A. *PNAS.* 2018;115:10523–8. <https://doi.org/10.1073/pnas.1803398115>.
 68. Slatnar A, Mikulic-Petkovsek M, Verberic R, Stampar F, Schmitzer V. Anthocyanin and chlorophyll content during poinsettia bract development. *Sci Hortic.* 2013;150:142–5. <https://doi.org/10.1016/j.scienta.2012.10.014>.
 69. Stewart RN, Asen S, Massie DR, Norris KH. The anthocyanin and flavonol composition of three families of poinsettia color sports. *J Hered.* 1980;71:140–2. <https://doi.org/10.1093/oxfordjournals.jhered.a109329>.
 70. Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, et al. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 2007;50:660–77. <https://doi.org/10.1111/j.1365-313X.2007.03078.x>.
 71. Tai D, Tian J, Zhang J, Song T, Yao Y. A *Malus crabapple* chalcone synthase gene, McCHS, regulates red petal color and flavonoid biosynthesis. *PLoS One.* 2014;9:e110570. <https://doi.org/10.1371/journal.pone.0110570>.
 72. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 2008;54:733–49. <https://doi.org/10.1111/j.1365-313X.2008.03447.x>.
 73. Thomas B. Light signals and flowering. *J Exp Bot.* 2006;57:3387–93. <https://doi.org/10.1093/jxb/erl071>.
 74. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 2017;45:W122–9. <https://doi.org/10.1093/nar/gkx382>.
 75. Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, et al. Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* 2005;42:218–35. <https://doi.org/10.1111/j.1365-313X.2005.02371.x>.
 76. Trejo L, Feria Arroyo TP, Olsen KM, Eguarte LE, Arroyo B, Gruhn JA, et al. Poinsettia's wild ancestor in the Mexican dry tropics: historical, genetic, and environmental evidence. *Am J Bot.* 2012;99:1146–57. <https://doi.org/10.3732/ajb.1200072>.
 77. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 2017;45:D158–69. <https://doi.org/10.1093/nar/gkw1099>.
 78. Valdés A, Ibáñez C, Simó C, García-Cañas V. Recent transcriptomics advances and emerging applications in food science. *TrAC Trends Anal Chem.* 2013;52:142–54. <https://doi.org/10.1016/j.trac.2013.06.014>.
 79. Vimolmangkang S, Han Y, Wei G, Korban SS. An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol.* 2013;13:176. <https://doi.org/10.1186/1471-2229-13-176>.
 80. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, et al. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun.* 2016;7:11708. <https://doi.org/10.1038/ncomms11708>.
 81. Wang C, Zhi S, Liu C, Xu F, Zhao A, Wang X, et al. Isolation and characterization of a novel chalcone synthase gene family from mulberry. *Plant Physiol Biochem.* 2017;115:107–18. <https://doi.org/10.1016/j.plaphy.2017.03.014>.
 82. Wang Z, Yu Q, Shen W, El Mohtar CA, Zhao X, Gmitter FG. Functional study of CHS gene family members in citrus revealed a novel CHS gene affecting the production of flavonoids. *BMC Plant Biol.* 2018;18:189. <https://doi.org/10.1186/s12870-018-1418-y>.
 83. Waterhouse RM, Seppay M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 2017. <https://doi.org/10.1093/molbev/msx319>.
 84. Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang X-J, et al. Comprehensive comparison of Pacific biosciences and Oxford Nanopore technologies and their applications to transcriptome analysis. *F1000 Res.* 2017;6:100. <https://doi.org/10.12688/f1000research.10571.2>.
 85. Wen C-H, Chu F-H. A R2R3-MYB gene LfMYB113 is responsible for autumn leaf coloration in Formosan sweet gum (*Liquidambar formosana* Hance). *Plant Cell Physiol.* 2017;58:508–21. <https://doi.org/10.1093/pcp/pcw228>.
 86. Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 2001;126:485–93. <https://doi.org/10.1104/pp.126.2.485>.
 87. Xu Q, Zhu J, Zhao S, Hou Y, Li F, Tai Y, et al. Transcriptome profiling using single-molecule direct RNA sequencing approach for in-depth understanding of genes in secondary metabolism pathways of *Camellia sinensis*. *Front Plant Sci.* 2017;8:1205. <https://doi.org/10.3389/fpls.2017.01205>.
 88. Yahya M, Ali S, Davidovich-Rikanati R, Ibdah M, Shachtier A, Eyal Y, et al. Characterization of three chalcone synthase-like genes from apple (*Malus x domestica* Borkh.). *Phytochemistry.* 2017;140:125–33. <https://doi.org/10.1016/j.phytochem.2017.04.022>.
 89. Yamagishi M, Uchiyama H, Handa T. Floral pigmentation pattern in oriental hybrid lily (*Lilium* spp.) cultivar 'dizzy' is caused by transcriptional regulation of anthocyanin biosynthesis genes. *J Plant Physiol.* 2018;228:85–91. <https://doi.org/10.1016/j.jplph.2018.05.008>.
 90. Yan K, Wei Q, Feng R, Zhou W, Chen F. Transcriptome analysis of *Cinnamomum longepaniculatum* by high-throughput sequencing. *Electron J Biotechnol.* 2017;28:58–66. <https://doi.org/10.1016/j.ejbt.2017.05.006>.
 91. Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, et al. Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.* 2012;69:154–67. <https://doi.org/10.1111/j.1365-313X.2011.04779.x>.
 92. Yoshihara N, Imayama T, Fukuchi-Mizutani M, Okuhara H, Tanaka Y, Ino I, et al. cDNA cloning and characterization of UDP-glucose: Anthocyanidin 3-O-glucosyltransferase in *Iris hollandica*. *Plant Sci.* 2005;169:496–501. <https://doi.org/10.1016/j.plantsci.2005.04.007>.
 93. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 2017;45:D744–9. <https://doi.org/10.1093/nar/gkx1119>.
 94. Zhang L, He L-L, Fu Q-T, Xu Z-F. Selection of reliable reference genes for gene expression studies in the biofuel plant *Jatropha curcas* using real-time quantitative PCR. *Int J Mol Sci.* 2013;14(12):24338–54. <https://doi.org/10.3390/ijms141224338>.
 95. Zhang KM, Wang JW, Guo ML, Du WL, Wu RH, Wang X. Short-day signals are crucial for the induction of anthocyanin biosynthesis in *Begonia semperflorens* under low temperature condition. *J Plant Physiol.* 2016;204:1–7. <https://doi.org/10.1016/j.jplph.2016.06.021>.
 96. Zhang L, Xu B, Wu T, Yang Y, Fan L, Wen M, et al. Transcriptomic profiling of two Pak Choi varieties with contrasting anthocyanin contents provides an insight into structural and regulatory genes in anthocyanin biosynthetic pathway. *BMC Genomics.* 2017;18:288. <https://doi.org/10.1186/s12864-017-3677-7>.
 97. Zhao J. Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.* 2015;20:576–85. <https://doi.org/10.1016/j.tplants.2015.06.007>.
 98. Zhao J, Dixon RA. The 'ins' and 'outs' of flavonoid transport. *Trends Plant Sci.* 2010;15:72–80. <https://doi.org/10.1016/j.tplants.2009.11.006>.
 99. Zhao ZC, Hu GB, Hu FC, Wang HC, Yang ZY, Lai B. The UDP glucose: flavonoid-3-O-glucosyltransferase (UFGT) gene regulates anthocyanin biosynthesis in litchi (*Litchi chinensis* Sonn.) during fruit coloration. *Mol Biol Rep.* 2012;39:6409–15. <https://doi.org/10.1007/s11033-011-1303-3>.
 100. Zhou B, Wang Y, Zhan Y, Li Y, Kawabata S. Chalcone synthase family genes have redundant roles in anthocyanin biosynthesis and in response to blue/UV-A light in turnip (*Brassica rapa*; Brassicaceae). *Am J Bot.* 2013;100:2458–67. <https://doi.org/10.3732/ajb.1300305>.
 101. Zhou X, Li J, Zhu Y, Ni S, Chen J, Feng X, et al. De novo assembly of the *Camellia nitidissima* Transcriptome reveals key genes of flower pigment biosynthesis. *Front Plant Sci.* 2017;8:1545. <https://doi.org/10.3389/fpls.2017.01545>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.