Molecular studies on the anthocyanin pathways in red- and white-flowering poinsettias (*Euphorbia pulcherrima* Willd. Ex. Klotzsch)

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Vinicius Vilperte, M. Sc.

2021

Referent: Prof. Dr. rer. nat. Thomas Debener

Korreferentin: Prof. Dr. rer. nat. Helge Küster

Tag der Promotion: 15.06.2021

Abstract

Poinsettia is a popular and important ornamental crop, mostly during the Christmas season. Its bract colouration ranges from pink/red to creamy/white shades, with nearly all white varieties being obtained through mutation breeding (γ- or X-ray mutagenesis) of red varieties. The appearance of acyanic poinsettia varieties is referred to as the '*white paradox'* since the expression of most structural genes and the related enzyme activities involved in the formation of red anthocyanin pigments are present.

To unravel the genetic factors responsible for the '*white paradox'*, we firstly assembled and functionally annotated a hybrid *de novo* bract transcriptome for the species, as well as characterized the main transcriptional differences between red- and white-bracted poinsettia varieties during bract development. The differential gene expression analysis pointed out a *glutathione S-transferase* (*GST*) gene as a putative candidate for the appearance of white genotypes. The poinsettia *GST* (named *Bract1*) is an active gene involved in the expression of anthocyanins in bracts and it presents a high phylogenetic similarity to known anthocyanin-related GSTs. We identified a 4 bp deletion in a short repeat within the coding region of *Bract1*, which is the most likely cause of many mutations that lead to a white bract colour. Moreover, overexpression of the *Bract1* wild-type allele in Arabidopsis *tt19* mutants restored the anthocyanin phenotype, while the *Bract1* mutated allele showed to be non-functional.

In poinsettia mutation breeding, not all red varieties can produce white sports through radiation; thus, they are distinguished into 'heterozygous' and 'homozygous' for the colouration locus according to their ability to generate white sports. The *Bract1* polymorphism between wild-type and mutant alleles co-segregates with the phenotype in progeny from heterozygous red and white parents, thus confirming that *Bract1* is linked to the colour trait. Based on a PCR assay for *Bract1*, we identified low-frequency heterozygous mutations arising from homozygous genotypes. By developing a multiplex sequencing approach, we were able to detect the presence of 1 mutated allele in a pool of 50 nonmutated copies. In conclusion, we identified a short repeat mutation in *Bract1* as the most likely cause for the '*white paradox*', which might also serve as a reference for the study of other repeat-containing structural genes as potential mutational hot spots in plant genomes. Moreover, we developed a multiplex approach that may enable an expansion of the genetic resources available for the development of new varieties.

Keywords: Hybrid transcriptome, mutation breeding, anthocyanin, *Bract1*.

Zusammenfassung

Poinsettien gehören vor allem in der Weihnachtszeit zu den beliebtesten und wichtigsten Zierpflanzen. Die Färbung der Brakteen reicht von rosa/rot bis creme/weiß, wobei fast alle weißen Sorten durch Mutationszüchtung aus roten Sorten entstanden sind. Das Auftreten von anthozyanfreien Poinsettien wird als "weißes Paradoxon" bezeichnet, weil die Expression der meisten Strukturgene und die damit verbundenen Enzymaktivitäten, die an der Bildung der roten Anthocyanpigmente beteiligt sind, nachweisbar sind.

Um die genetischen Faktoren zu charakterisieren, die für das "weiße Paradoxon" verantwortlich sind, wurde ein *de novo* Transkriptom auf Basis verschiedener Brakteen erstellt und nach den wichtigsten transkriptionellen Unterschieden zwischen roten und weißen Sorten während der Brakteenentwicklung gesucht. Durch differentielle Genexpressionsanalyse konnte ein *glutathion-S-transferase* (*GST*)-Gen als wahrscheinlichstes Kandidatengen für das Auftreten der weißen Genotypen identifiziert werden. In Poinsettien ist *GST* (genannt *Bract1*) ein aktives Gen, das an der Expression von Anthocyanen in den Brakteen beteiligt ist. Im kodierenden Bereich von *Bract1* konnte eine 4 bp Deletion in einer kurzen Wiederholung einer Basenabfolge identifiziert werden, die vermutlich die Ursache für viele Mutationen ist, die zu weißen Brakteen führen. Eine Überexpression des *Bract1*-Wildtyp-Allels in Arabidopsis *tt19*-Mutanten stellte den Anthocyan-Phänotyp wieder her, während das mutierte *Bract1*-Allel nicht funktional war. Nicht alle roten Poinsettien-Sorten haben die gleiche Fähigkeit durch Bestrahlung weiße Nachkommen erzeugen; daher werden sie in 'heterozygot' und 'homozygot' für den Locus, der für die Färbung verantwortlich, unterschieden. Der *Bract1*-Polymorphismus zwischen Wildtyp und Mutanten kosegregiert mit dem Phänotyp in den Nachkommen von heterozygot roten und weißen Eltern, was bestätigt, dass *Bract1* an das Farbmerkmal gekoppelt ist. Mit einem für *Bract1* entwickelten PCR-Assay konnten wir eine geringe Anzahl an heterozygoten Nachkommen, die aus homozygoten Genotypen hervorgingen, nachweisen. Mit der Entwicklung eines Multiplex-equenzierungsansatzes ist es auch möglich eine einzelne Kopie des mutierten Allels in einem Pool von 50 nicht mutierten Kopien nachzuweisen. Dieser Ansatz könnte dazu genutzt werden, um die verfügbaren genetischen Ressourcen für die Entwicklung neuer Sorten zu erweitern. Des Weiteren könnte die identifizierte Mutation in *Bract1* als Basis für die Analyse anderer Strukturgene als potenzielle Mutations-Hotspots in Pflanzengenomen dienen.

Schlagwörter: Hybrid-Transkriptom, Mutationszüchtung, Anthocyanin, *Bract1*.

Content

Abstract ..... **I**

Zusammenfassung ..... **II**

Content ..... **III**

Abbreviations ..... **V**

1 General Introduction **1**

1.1 Poinsettia 1

1.1.1 Taxonomy, geographic distribution, and morphology 1

1.1.2 Genetics and molecular aspects of poinsettia 4

1.1.3 Breeding history – from Mexico to the US 4

1.2 Plant pigments: A colourful pathway 5

1.2.1 Anthocyanin biosynthesis and regulation 7

1.2.2 Mechanisms of anthocyanin transport 10

1.2.3 Glutathione S-transferase as anthocyanin transporter 11

1.2.4 Colour formation in poinsettia: the *'white paradox'* 12

1.3 Transcriptomics 13

1.3.1 Transcriptome assembly and its challenges 14

1.3.2 The power of Single-Molecule Sequencing for transcriptome assembly 15

1.4 Mutation breeding 17

1.4.1 Mutation breeding in ornamentals 18

1.4.2 Microsatellites as a source of variation 20

1.5 Objectives 23

2 Hybrid de novo transcriptome assembly of poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) bracts **24**

2.1 Supporting Information 43

3 A highly mutable GST is essential for bract colouration in *Euphorbia pulcherrima* Willd. Ex Klotsch **49**

3.1 Supporting information 66

4 Development of a multiplex amplicon-sequencing assay to detect low-frequency mutations in poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) breeding programmes 71

4.1 Supporting information 83

5 General discussion 84

5.1 The transcriptome of poinsettia bracts 84

5.2 Improving and expanding the knowledge on the genetic basis of bract colouration in poinsettia 86

5.3 *Bract1* is a functional anthocyanin transporter in poinsettia bracts 88

5.4 Development of an assay for mutation detection in poinsettia breeding programmes 90

5.5 Prospect biotechnological applications in poinsettia breeding 92

5.6 Conclusion 94

References 96

Electronic appendix 126

Curriculum vitae 128

Acknowledgement 132

Abbreviations

A	
ABC	ATP binding cassette
AFLP	Amplified fragment length polymorphism
ANS	Anthocyanidin synthase
ANR	Anthocyanidin reductase
B	
Bp	Base pair
bHLH	Basic helix-loop-helix
BZ2	Bronze 2
C	
CHI	Chalcone isomerase
CHS	Chalcone synthase
CAGE	Cap analysis gene expression
cDNA	Complementary DNA
CCS	Circular consensus sequence
D	
DNA	Deoxyribonucleic acid
DHK	Dihydrokaempferol
DHM	Dihydromyricetin
DDRT-PCR	Differential-Display Reverse Transcription-PCR
DHAR	Dehydroascorbate reductase
DSB	Double-strand break
DFR	Dihydroflavonol 4-reductase
E	
EBG	Early biosynthetic genes
ER	Endoplasmic reticulum
EST	Expressed sequence tags
EF1A/B	Translation elongation factor 1 alpha/beta
EMS	Ethyl methanesulfonate
F	
F3H	Flavanone 3-hydroxylase
F3’H	Flavonoid 3′-hydroxylase
F3’5’H	Flavonoid 3′,5′-hydroxylase
FLS	Flavonol synthase
G	
GST	Glutathione S-transferase
GHR	Glutathionyl hydroquinone reductase
H	
HR	Homologous recombination

I	
Iso-Seq	Isoform sequencing

L	
LBG	Late biosynthetic genes
LAR	Leucoanthocyanidin reductase
LDOX	Leucoanthocyanidin dioxygenase
LT	Ligandin Transporter

M	
MBW	MYB-bHLH-WD40 complex
MATE	Multidrug and toxic compound extrusion
MPSS	Massively Parallel Signature Sequencing
MMEJ	Microhomology-mediated end joining
MMS	Methyl methanesulfonate
mPGES-2	Microsomal prostaglandin E-synthase type 2
MNU	N-methyl-N-nitrosourea

N	
NGS	Next-generation sequencing
ncRNA	Non-coding RNA
NHEJ	Non-homologous end-joining

O	
ONT	Oxford Nanopore

P	
PnMV	Poinsettia Mosaic Virus
PCR	Polymerase chain reaction
PacBio	Pacific Biosciences

R	
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
ROS	Reactive oxygen species
RAPD	Random Amplified Polymorphic DNA

S	
SMS	Single molecular sequencing
SMRT	Single-molecule real-time
SAGE	Serial Analysis of Gene Expression
sRNA	Small RNA
SNP	Single nucleotide polymorphism
SSR	Single Sequence Repeats
STR	Short tandem repeats
SSA	Single-strand annealing

T

TGS	Third Generation Sequencing
tt19	Transparent testa 19
TCHQD	tetra-chloro hydroquinone dehalogenase

V

VT	Vesicular Transport

1 General Introduction

1.1 Poinsettia

Poinsettia, also known as nochebuena, Weihnachtsstern, star of Bethlehem, or Christmas star (Trejo et al., 2012), is an important ornamental crop, especially due to its association with Christmas time in North America, Europe, and Asia. Annual sales in the USA in 2018 reached nearly 150 million dollars, accounting for a 6% increase in revenue compared to 2015, with around 34 million potted poinsettias sold in the same year (USDA, 2018). Meanwhile, in Germany, total consumption is estimated at around 32 million potted poinsettias in 2018 (van der Ploeg, 2018). Its ornamental value relies on the intense coloured bracts, such as red, white, pink, yellow, or even bicoloured, scattered or marbled colourations. Nonetheless, poinsettia breeding still focuses on red and white coloured varieties due to a higher acceptance by the consumers. In 2018, around 80% of poinsettias grown in Germany had red bract, 11% had white bracts, and 9% had pink or bicolored/scattered coloured bracts (van der Ploeg, 2018).

1.1.1 Taxonomy, geographic distribution, and morphology

Euphorbia pulcherrima Willd. ex Klotsch belongs to the genus Euphorbia L. (family: Euphorbiaceae) (Table 1), which comprises more than 2,000 species and represents one of the largest genera within angiosperms (Govaerts et al., 2000; Zimmermann et al., 2010). The rich morphological variability of species within the genus, e.g. inflorescences and seeds, creates a challenge to its taxonomic classification but also distinguishes it as a promising genus to investigate evolutionary questions related to plant morphology and diversification (Horn et al., 2012). One of these characteristics that drove the diversification of the clade is the unique reproductive structure present in *Euphorbia*, the cyathium, which resembles a flower but is widely interpreted as a condensed inflorescence (Prenner et al., 2008, 2010).

Within the genus, four subgenera are recognized to date based on phylogenetic studies: i) *Rhizanthium* (Clade A) – approximately 200spp. with primarily African distribution; ii) *Esula* (Clade B) – around 500spp. with herbaceous species distributed in the North Hemisphere from the eastern Mediterranean through Asia, and woody and succulent species in Africa; iii) *Euphorbia* (Clade C) – around 700spp. and the most diverse subgenus. It includes xerophytes in Africa, India, South and Central America and

Australia; and iv) *Chamaesyce* (Clade D) – around 600spp.containing the majority of the New World Euphorbia species, with a large of them being C_4 photosynthetic herbs and subshrubs, as well as the newly classified ‘‘*Agaloma* alliance’’, including *E. pulcherrima*, which is native to the dry tropical forests of Mexico and Guatemala (Steinmann and Porter, 2002; Bruyns et al., 2006; Park and Jansen, 2007; Zimmermann et al., 2010; Horn et al., 2012; Trejo et al., 2012; Pérez-Nicolás et al., 2020).

Table 1. Taxonomic classification of poinsettias (GBIF Secretariat, 2019).

Kingdom	Plantae
Subkingdom	Viridiplantae
Phylum	Tracheophyta
Class	Magnoliopsida
Order	Malphigiales
Family	Euphorbiaceae
Genus	*Euphorbia*
Subgenus	*Chamaesyce*

The wild poinsettia is a small deciduous shrub occurring in mid-elevation tropical forests of the Mexican tropical Pacific slope with hairless cane-like stems reaching more than 6 meters and unbranched and distantly spaced leaves (Trejo et al., 2012, 2018). Leaves are long petiolated, 10-15 cm long, relatively thin and with pointed tips. Male and female flowers are grouped in the cyathia, which bears a two-lipped yellow gland, and they develop through the differentiation of bract axillary buds (Figure 1A-D) (Kannangara and Hansson, 1998; Bercu, 2016). Flowering induction occurs under short-day conditions, which is accompanied by the development and colouration of bracts (Kannangara and Hansson, 1998; Vilperte et al., 2019), a modified or specialized leaf, often associated with a reproductive structure such as a flower or inflorescence. Breeding efforts have selected modern cultivars with greater bracts, shortened internodes, compact and more rounded outlines with a bushy habit (Figure 1E-H) (Taylor et al., 2011; Kobayashi, 2012; Trejo et al., 2018).

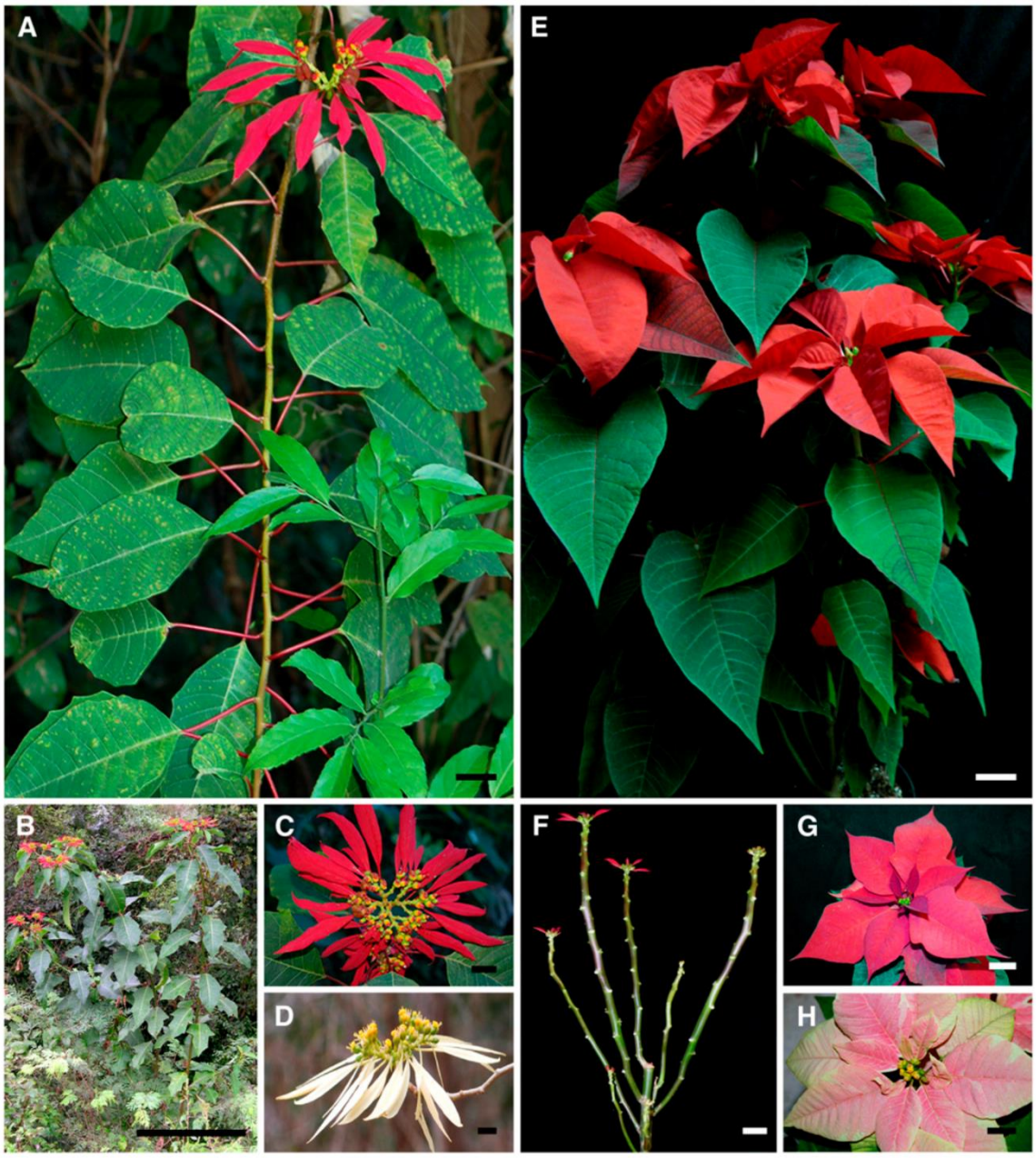


Figure 1. Phenotypical differences between wild and domesticated plants of *Euphorbia pulcherrima*. (A) Wild plant in Oaxaca, Mexico, showing the small, sparse bracts, tall habit, and long internodes. (B) Growth habit of a wild plant showing the tall stature, commonly reaching 3–4 m, and the sparse branching. (C) Inflorescence with the abundant flowers that characterize wild plants. (D) Inflorescence with white bracts, an extremely uncommon variant. (E) Growth habit of a domesticated plant, showing the wide, abundant bracts and short internodes. (F) Highly branched domesticated plant. (G) Dense and abundant bracts, few flowers, and no fruits of domesticated plants. (H) Some domesticated varieties are pink, white, plum-coloured, or even mottled. Scale bars in A, C–H = 3 cm; in B = 1m (Adapted from Trejo et al., (2012)).

1.1.2 Genetics and molecular aspects of poinsettia

Modern poinsettia varieties are mainly diploid with a basic chromosome number of 14 (2n=2x=28), but also triploid (2n=3x=42) and tetraploid (2n=4x=56) varieties have been used for breeding in the past (Moyer, 1934; Ewart and Walker, 1960; Bempong and Sink, 1968b, 1968a). Genetic and molecular studies on the species are scarce. Isoenzymes were the first markers used for cultivar differentiation (Werner and Sink, 1977), with DNA based markers, i.e. RAPD markers (Ling et al., 1997) and AFLP markers (Parks and Moyer, 2004), only available decades later. Besides, the genetic diversity of the species in its native regions, as well as the introgression of wild poinsettia in garden plants and modern cultivars, has been analysed using plastid and nuclear DNA markers (Trejo et al., 2012, 2019; Trejo-Hernández et al., 2015). Moreover, a transcriptomic study has identified a set of genes involved in anthocyanin pigmentation in poinsettia bracts, as well as generated resourceful information on a molecular level for the species (Gu et al., 2018).

Genetic transformation of poinsettia is still not exploited. Protoplast transformation showed to be a suitable and high efficient method for transient expression (Pitzschke and Persak, 2012). *Agrobacterium tumefaciens*-mediated stable transformation was applied to induce resistance to Poinsettia Mosaic Virus (PnMV) by using virus-derived hairpin (hp) RNA gene constructs to induce RNA silencing (Clarke et al., 2008). Besides, *A. tumefaciens*-mediated transformation was also used to overexpress the *Arabidopsis SHORT INTERNODE* (*AtSHI*) gene and resulted in plants with reduced plant height and internode lengths (Islam et al., 2013).

1.1.3 Breeding history – from Mexico to the US

The species common name, poinsettia, honours the American ambassador in Mexico in the 1820's Joel Roberts Poinsett, who was the first to collect living material of *E. pulcherrima* from Mexico and sent it over to Philadelphia, US (Lack, 2011). The plants were cultivated in the botanical garden of the Bartram family and, at that time, they were taxonomically classified as *Pleuradena coccinea* Raf., but also known as *Euphorbia poinseti* (Lack, 2011). Next, Robert Bust sent cuttings from Philadelphia to the Royal Botanic Garden in Edinburgh and from there to the Royal Botanic Garden in Schöneberg near Berlin, where it reached the German botanist Karl Willdenow who named it *Euphorbia pulcherrima* in 1834 (Lack, 2011; Taylor et al., 2011). Trejo et al. (2012)

showed through species distribution modelling and genetic analysis that modern poinsettia varieties are derived from the region of Guerrero, Mexico, where Poinsett first collected them.

The genetic base of modern breeding varieties in poinsettia is relatively narrow, mainly due to the limited gene flow of the species at their original collection sites (Trejo et al., 2012). Commercial cultivars originated by hybridisation or somatic mutation from ’Oak Leaf’ and ’Ecke White’, with all of them belonging to ’Annette Hegg’ or ’Paul Mikkelsen’ cultivar groups (Werner and Sink, 1977; Stewart et al., 1979). Cultivars were divided into two morphotypes: i) free branching, characterized by weak apical dominance, many axillary shoots and relatively small stem diameters; and ii) restricted branching, characterized by strong apical dominance, few axillary shoots and large stem diameters (Dole and Wilkins, 1991).

To eliminate potential pathogens (e.g. Poinsettia Mosaic Virus - PnMV), poinsettias are subjected to heat treatment, meristem culture or somatic embryogenesis, which leads to a loss of the free branching ability. However, the trait can be restored by grafting these plants onto free branching cultivars (Lee, 2000). The success of this approach is due to the discovery of a phytoplasma present in the free branching cultivars, and its ability to be transmitted by grafting to restricted branching ones, thus opening up possibilities for poinsettia breeding by allowing the use of diverse and disease-free cultivars (Lee et al., 1997). An infectious agent related to peach X disease and spirea stunt was identified as the cause for the phytoplasma, which showed no pathogenicity in poinsettia and thus became a standard procedure to induce branching in new poinsettia seedlings by grafting (Lee et al., 1997; Taylor et al., 2011).

1.2 Plant pigments: A colourful pathway

Plant pigmentation is of outstanding importance in plant ecology due to its ability to attract pollinators and seed-dispersing organisms (Kevan and Baker, 1983; Mol et al., 1998; Grotewold, 2006a), as well as in plant-environment interactions, such as light protection and antioxidants, chelating agents for metals, and protection against biotic and abiotic stresses (Chalker-Scott, 1999; Harvaux and Kloppstech, 2001; Ahmed et al., 2014; Landi et al., 2015; Jordan, 2018). In ornamental plants, flower colour is a key feature, determining the quality and ornamental merits, as well as influencing its commercial value. Therefore, understanding the mechanisms underlying colour biosynthesis,

regulation and transport is essential for the cultivation and breeding of ornamental plants (Zhao and Tao, 2015).

The colouration of flowers is due to the accumulation of three chemically distinct pigments: flavonoids, carotenoids, and betalains. The first two classes are widely distributed in the plant kingdom, whereas betalains are found exclusively in one group of angiosperms, the order Caryophyllales (e.g. beetroot, *Amaranthus* and portulaca) (Mol et al., 1998; Grotewold, 2006a; Tanaka et al., 2008). Betalains comprise the red-violet betacyanins and yellow-orange betaxanthins and, although much scientific interest is driven by these pigments, its biosynthesis, modulation, and regulatory pathways have not been fully elucidated (Brockington et al., 2011; Hatlestad et al., 2012, 2015; Chung et al., 2015; Davies, 2015). Interestingly, betalains and anthocyanin seem to be exclusive pathways and have never been found together in any plant (Stafford, 1994; Sakuta, 2014).

Beyond their long recognized function as plant pigments ranging from red to yellow hues (Grotewold, 2006a), carotenoids are essential for light capture, photoprotection, and stabilization of the photosynthetic apparatus in plants (Hashimoto et al., 2016; Stanley and Yuan, 2019), as well as providing precursors for the biosynthesis of phytohormones (Nambara and Marion-Poll, 2005; Al-Babili and Bouwmeester, 2015) and acting as signalling molecules to plant development (Tian, 2015). The carotenoid metabolism, as well as their regulation and biosynthesis, has been extensively studied due to their pivotal roles in plants and human nutrition (Ruiz-Sola and Rodríguez-Concepción, 2012; Fiedor and Burda, 2014; Nisar et al., 2015; Yuan et al., 2015; Stanley and Yuan, 2019).

Flavonoids are widely distributed secondary metabolites in plants estimated to comprise over 8000 metabolites and they can be classified into six major subgroups: flavones, flavonols, flavanones, flavanols, anthocyanidins and isoflavones, therefore leading to a wide spectrum of colours, from pale yellow to blue-purple (Zhao and Tao, 2015; Tohge et al., 2017). Besides their known main function as plant pigments (Mol et al., 1998; Koes et al., 2005), flavonoids have important roles in photoprotection (Agati and Tattini, 2010), auxin transport (Peer and Murphy, 2007), allelopathy (Bais et al., 2006), modulation of reactive oxygen species (ROS) levels (Taylor and Grotewold, 2005), as well as defence against herbivores and pathogens (Peters and Constabel, 2002; Torregrosa et al., 2004).

1.2.1 Anthocyanin biosynthesis and regulation

Anthocyanin pigments are responsible for the majority of the orange, red, purple, and blue colours of flowers, and their biosynthesis, which derives from a branch of the flavonoid pathway, has been a research hotspot in the field of plant secondary metabolism (Tanaka et al., 2008; Cheynier et al., 2013; Saito et al., 2013; Zhao and Tao, 2015). Flavonoid synthesis occurs at the convergence of the shikimate and acetate pathways, subsequently through the phenylpropanoid pathway, transforming phenylalanine into 4-coumaroyl-CoA, which finally enters the flavonoid biosynthesis pathway with *chalcone synthase* (*CHS*) providing the precursor for all classes of flavonoids (Falcone Ferreyra et al., 2012; Tohge et al., 2017).

The anthocyanin specific pathway is depicted in Figure 2. Briefly, it starts with *CHS* mediated synthesis of naringenin chalcone from 4-coumaroyl-CoA and malonyl-CoA. *Chalcone isomerase* (*CHI*) then isomerizes naringenin chalcone to naringenin. *Flavanone 3-hydroxylase* (*F3H*) converts naringenin into dihydrokaempferol (DHK) that can be further hydroxylated by *flavonoid 3'-hydroxylase* (*F3'H*) or *flavonoid 3',5'-hydroxylase* (*F3'5'H*) into dihydroquercetin (DHK) or dihydromyricetin (DHM), respectively. Next, the three dihydroflavonols are converted into colourless leucoanthocyanidins by *dihydroflavonol 4-reductase* (*DFR*) and subsequently to coloured anthocyanidins by *anthocyanidin synthase* (*ANS*). Finally, various members of the glycosyltransferase enzyme family attach sugar molecules to anthocyanidins, e.g. *flavonoid 3-O-glucosyltransferase* (*UFGT*). In addition to glucose, anthocyanins containing rhamnose and other sugars at the 5, 3', and 7 positions are also found in different plants, as well as acylated ones by a variety of organic acids through the action of anthocyanin acyltransferases (Grotewold, 2006a; Sasaki et al., 2014; Liu et al., 2018b)

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* (Petroni and Tonelli, 2011). 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 (Dubos et al., 2010), which in turn can be regulated by a cascade of signalling pathways (Bulgakov et al., 2017). *MYB* factors alone have been demonstrated to positively regulate the expression of EBGs in several species (Adato et al., 2009; Czemmel et al., 2009; Vimolmangkang et al., 2013). In Arabidopsis, knockout

of *MYB12* led to a clear reduction of *CHS* and *FLS* expressions, while the expression of *CHI*, *F3H*, *DFR*, and *F3'H* remained unchanged (Mehrtens et al., 2005). Other *MYB* classes, such as the *R2R3-MYB* factors *PAP1*, *PAP2*, *MYB113*, *MYB114*, are involved in MBW complexes and regulate the expression of LBGs (Baudry et al., 2004; Espley et al., 2007; Butelli et al., 2008; Gonzalez et al., 2008).

Despite being the main transcriptional regulators of the flavonoid pathway together with *MYB* and *WD40* (Hichri et al., 2011), *bHLH* transcription factors, also known as MYCs, regulate several biological processes in plants, such as the development of floral organs (Heisler et al., 2001; Sorensen et al., 2003), photomorphogenesis (Ortigosa et al., 2020) and hormonal response (Dombrecht et al., 2007; Kazan and Manners, 2013; Chakraborty et al., 2019). Flavonoid-related bHLHs were first reported in maize (Chandler et al., 1989), and their importance in acting within the MBW complex has been observed in a range of species (Baudry et al., 2004; Shan et al., 2019; Wang et al., 2019c). Similarly, WD40 proteins are also involved in a range of roles in plants, e.g. cell division, signal transduction, RNA processing and regulation of transcription (van Nocker and Ludwig, 2003), and the play crucial roles in the flavonoid MBW regulatory complex (Yao et al., 2017; Liu et al., 2018a; Zhao et al., 2019).

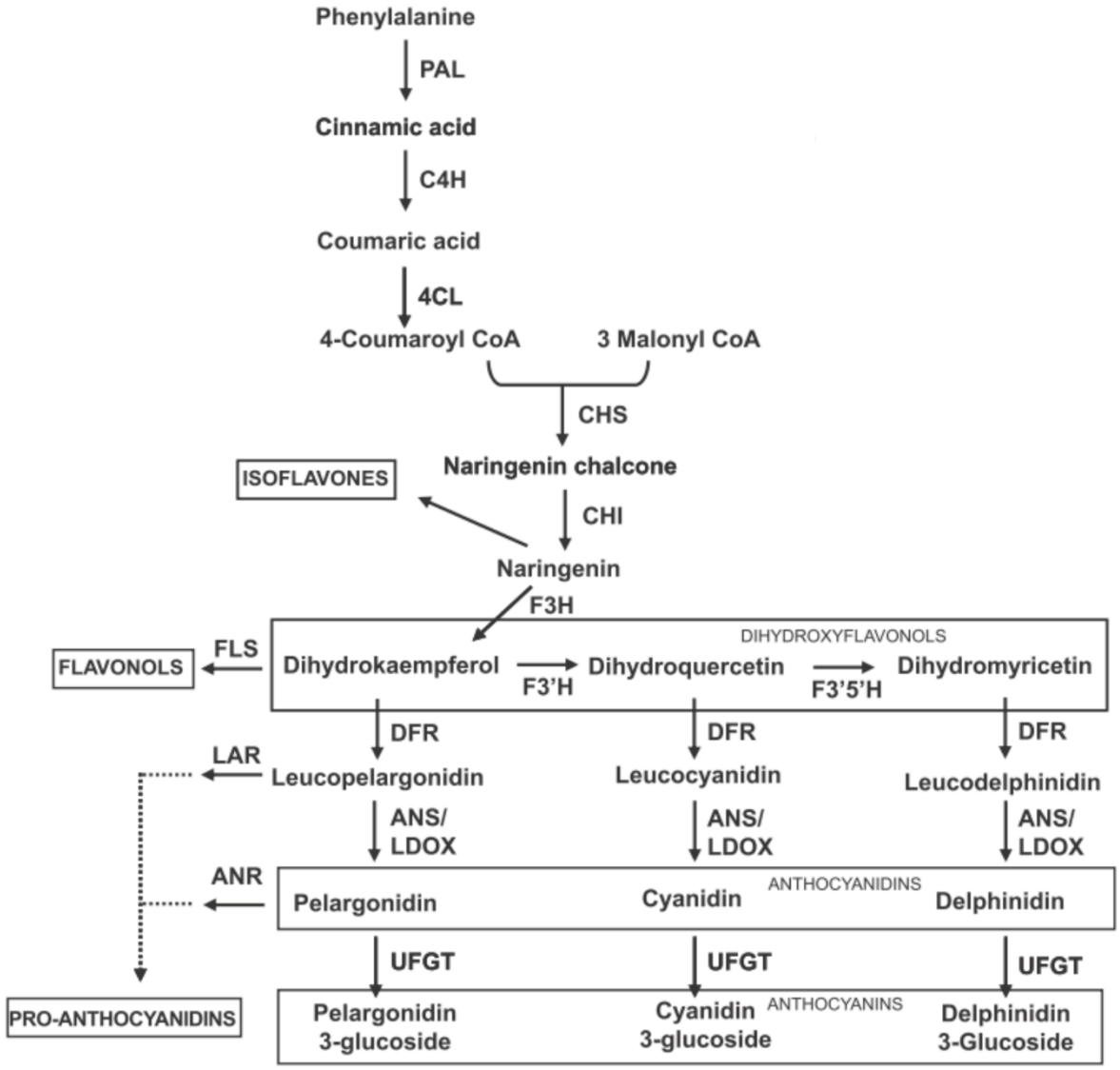


Figure 2. Simplified schematic representation of the anthocyanin biosynthetic pathway and other subgroups of flavonoids. Enzyme names are abbreviated as follows: *PAL*, *phenylalanine ammonia-lyase*; *C4H*, *cinnamic acid 4-hydroxylase*; *4CL*, *4 coumarate CoA ligase*; *CHS*, *chalcone synthase*; *CHI*, *chalcone isomerase*; *F3H*, *flavanone 3-hydroxylase*; *F3'H*, *flavanone 3-hydroxylase*; *F3'5'H*, *flavanone 3'5'-hydroxylase*; *DFR*, *dihydroflavonol reductase*; *FLS*, *flavonol synthase*; *ANS/LDOX*, *anthocyanidin synthase/leucoanthocyanidin dioxygenase*; *UFGT*, *UDP-flavonoid glucosyltransferase*; *ANR*, *anthocyanidin reductase*; *LAR*, *Leucoanthocyanidin reductase* (Adapted from Petroni and Tonelli (2011)).

1.2.2 Mechanisms of anthocyanin transport

The synthesis of anthocyanins takes place on the cytoplasmic surface of the endoplasmic reticulum (ER), with probably all enzymes establishing one or more ER-associated multienzyme complexes (Winkel, 2004). Once synthesized, anthocyanin molecules need to be stored in the vacuole to prevent oxidation, as well as to provide the acidic vacuolar environment necessary for anthocyanins to function as pigments (Marrs et al., 1995; Verweij et al., 2008; Faraco et al., 2014; Bassham, 2015). Thus, two non-exclusive possible models for the transport of anthocyanins from the ER surface to the vacuole have been proposed and extensively studied: the Vesicle trafficking-mediated model (Vesicular Transport - VT) and the transporter-mediated model (Ligandin Transporter - LT) (Grotewold and Davies, 2008).

The vesicle trafficking-mediated model hypothesizes that anthocyanins first translocate into the ER lumen and are then transported in vesicles and/or membrane-bound organelles to the vacuole, a process analogous to the mechanism by which proteins are transported to the vacuole (Bassham, 2015; Chanoca et al., 2015). This model of action is supported by early observations of anthocyanins and other flavonoids accumulation in the cytoplasm in discrete structures, such as in sweet potato (Nozue and Yasuda, 1985) and maize (Grotewold et al., 1998), and later observed in other species, e.g. lisianthus (*Eustoma grandiflorum*) (Zhang et al., 2006a), *Brassica napus* (Hsieh and Huang, 2007), Arabidopsis (Poustka et al., 2007), and grapevine (Conn et al., 2010; Gomez et al., 2011). This model also involves an autophagy-like pathway, where anthocyanin-containing ER vesicles could be engulfed by autophagosomes and deposited inside the vacuole (Pourcel et al., 2010; Chanoca et al., 2015).

The second and most established model is the transporter-mediated model, where a *glutathione S-transferase* (*GST*) protein and tonoplast transporter proteins are involved in the transport of anthocyanins (Tanaka et al., 2008; Zhao, 2015). It was first proposed that the *GST* role was to form anthocyanin–glutathione conjugates, which would be recognized by a tonoplast-localized pump, and subsequently transported into vacuoles (Marrs et al., 1995). However, *GST* enzymatic activity is not required for the *GST*-dependent anthocyanin vacuolar sequestration, suggesting that the *GST* can bind to anthocyanin molecules to form a complex, thus escorting them from the ER to the vacuole preventing oxidation (Alfenito et al. 1998; Mueller et al., 2000; Larsen et al., 2003; Zhao and Dixon, 2010). Moreover, *Multidrug and Toxic Compound Extrusion*

(*MATE*) transporters, a kind of ABC-transporter proteins, participate in the uptake of cytoplasmic anthocyanins into the vacuole (Debeaujon et al., 2001; Goodman et al., 2004; Marinova et al., 2007; Gomez et al., 2009).

1.2.3 Glutathione S-transferase as anthocyanin transporter

Glutathione S-transferases (GSTs) represent a widespread multigenic enzyme family, abundantly found in animals, plants and even in some prokaryotes, and involved in multifunctional roles (Edwards et al., 2000; Lallement et al., 2014). Plant GSTs were initially identified as playing major roles in herbicide detoxification (Frear and Swanson, 1970; Dixon et al., 1998; Chronopoulou et al., 2017), but later described to be involved in responses to biotic and abiotic stresses (Roxas et al., 2000; Agrawal et al., 2002; Gullner et al., 2018), hormone responses and biosynthesis (Frova, 2006; Oakley, 2011), developmental changes (Moons, 2005) and detoxification of xenobiotics (Dixon et al., 1998, 2010; Cummins et al., 2011; Kumar and Trivedi, 2018).

The classification of GSTs is based on sequence conservation, genomic organization, physiochemical properties, among others (Edwards et al., 2000; Lallement et al., 2014; Islam et al., 2019). So far, 14 GSTs classes have been identified in plants: tau (U), phi (F), lambda (L), dehydroascorbate reductase (DHAR), theta (T), zeta (Z), eukaryotic translation elongation factor 1B-γ subunit (EF1Bγ), tetra-chloro hydroquinone dehalogenase (TCHQD), microsomal prostaglandin E-synthase type 2 (mPGES-2), glutathionyl hydroquinone reductase (GHR), metaxin, Ure2p, hemerythrin (H) and iota (I) (reviewed by (Lallement et al., 2014)). A large number of GSTs have been identified in plants species, such as 49 in *Capsella rubella* (He et al., 2016), 55 in Arabidopsis (Dixon and Edwards, 2010), 61 in Citrus (Licciardello et al., 2014) and 139 in *L. chinensis* (Hu et al., 2016).

The active role of GSTs in anthocyanin transportation has been demonstrated for several plant species (Hu et al., 2016; Pérez-Díaz et al., 2016; Jiang et al., 2019; Kou et al., 2019; Liu et al., 2019). The *Bronze-2* (*Bz2*) gene from maize was the first one reported to act as an anthocyanin transporter, which was demonstrated by the accumulation of anthocyanins in the cytoplasm in *bz2* mutants (Marrs et al., 1995). In petunia, the appearance of white mutants was attributed to a mutation on the *an9* gene, with the phenotype being complemented by the overexpression of *An9* and *Bz2* (Alfenito et al. 1998). A mutation on the *fl3* gene from carnation leads to a light pink phenotype, but with

a brighter phenotype being complemented by the petunia *An9* and maize *Bz2* (Larsen et al., 2003). The anthocyanin reduction of the Arabidopsis *transparent testa 19* mutant was complemented by the overexpression of *An9* (Kitamura et al., 2004). In peach, four alleles of a GST gene (*Riant*) were identified, with two of them containing frameshift mutations and not able to complement the Arabidopsis *tt19* phenotype. Varieties containing copies of the mutated alleles in a homozygous state showed flowers with white variegated phenotypes (Cheng et al., 2015). A Single Nucleotide Polymorphism (SNP) in the strawberry *reduced anthocyanins in petioles* (*RAP*) gene, leading to a premature stop codon, results in a mutant with green petioles and leaves. The non-functional *rap* gene was not able to complement the Arabidopsis *tt19*, while the wild-type *RAP* was successful (Luo et al., 2018).

1.2.4 Colour formation in poinsettia: the *'white paradox'*

The coloured bracts of poinsettias, varying from red, white, pink, yellow, or even bicoloured, scattered, or marbled colourations, determines the ornamental value of the species. In poinsettia, bract axillary buds differentiate into flowers (Kannangara and Hansson, 1998) under short-day conditions, which is accompanied by the development and colouration of bracts, thus indicating that the anthocyanin metabolism is regulated by photoperiodism (Jaakola and Hohtola, 2010). Green leaves and red bracts occur concomitantly in poinsettia plants and they accumulate different groups of pigments, i.e. chlorophylls and anthocyanins (Pomar and Ros Barceló, 2007; Moustaka et al., 2018). Several anthocyanin types have been identified in poinsettia bracts, with cyanidin and pelargonidin accounting for almost the complete total (Asen, 1958; Stewart et al., 1979, 1980; Slatnar et al., 2013; Nitarska et al., 2018).

Although the anthocyanin content of poinsettias has been previously reported, studies on the molecular mechanism involved in the anthocyanin pathway are scarce. A *Dihydroflavonol-4-reductase* (*DFR*) was suggested to promote the conversion of green leaves into red bracts (Gu et al., 2018). The orange colour of the poinsettia cultivar 'Harvest Orange' was related to a nonsense mutation in a *flavonoid 3'-hydroxylase* (*F3'H*) allele, thus leading to a higher accumulation of pelargonidin (Nitarska et al., 2018). The pink colouration in pink and marble bracts can be attributed to periclinal chimeric structures (Nielsen et al., 2003), while the reason for white/creamy colouration remains uncertain. Previous unpublished studies, together with our poinsettia transcriptome

(Vilperte et al., 2019), have observed the expression of most structural genes and the related enzyme activities involved in the formation of red anthocyanin pigments; thus, the appearance of acyanic (uncoloured) varieties is poinsettias is referred to as the '*white paradox*'.

1.3 Transcriptomics

The RNA was only separated from the DNA world in 1958 when Francis Crick formulated the hypothesis of the Central Dogma of Biology, where he proposed the idea that the genetic information is transcribed from DNA into RNA and further translated into proteins (Crick, 1958; Cobb, 2017). The transcriptome, defined as the complete set of DNA transcripts (i.e. RNAs) in a single cell and its quantities, for a specific stage of development or physiological condition, represents an important relation between the phenotype and the information encoded by DNA (Wang et al., 2009; Valdés et al., 2013). Transcriptomic studies became a promising field of studies in the post-genomic era (Lockhart and Winzeler, 2000), for reasons such as the ability to reflect spatiotemporal dynamics of gene expression, support to proteomics studies, and the possibility of studies regarding the structure and function of non-coding RNAs (ncRNAs) (Dong and Chen, 2013). The main goals of transcriptomic studies are: i) to identify all types of transcripts, including mRNAs, ncRNAs and small RNAs (sRNAs); ii) to determine gene structures, in terms of start sites, 5' and 3' ends, splicing patterns and other post-transcriptional modifications; and iii) to quantify changes in expression levels of each transcript (Wang et al., 2009).

Throughout the years, several techniques have been used to analyse gene expression, with the first approaches based on hybridization, such as cDNA microarrays and tilling genome arrays (Stolc et al., 2005; Draghici et al., 2006; Karakach et al., 2010; Ward et al., 2012). On the other hand, sequencing-based approaches, which can directly determine the cDNA sequence, have been intensively applied in transcriptomic studies. Several techniques were used in the last decades, such as Expressed Sequence Tags (ESTs) (Boguski et al., 1994; Blair et al., 2011), Differential Display Reverse Transcriptase Polymerase Chain Reaction (DDRT-PCR) (Liang and Pardee, 1992), Serial Analysis of Gene Expression (SAGE) (Velculescu et al., 1995), Cap Analysis Gene Expression (CAGE) (Shiraki et al., 2003) and Massively Parallel Signature Sequencing (MPSS) (Brenner et al., 2000). For many years, microarrays have been the main choice

for gene expression studies for organisms with sufficient genomic information, providing high-throughput, cost-effective and easily optimized solution to examine transcript expression (Malone and Oliver, 2011; Zhao et al., 2014; Hsieh et al., 2019). On the other hand, for non-model organisms, expression microarrays need to rely on cross-species hybridization (Vijay et al., 2013).

As an alternative, RNA sequencing (RNA-Seq) methods, defined as the direct sequencing of transcripts through Next Generation Sequencing (NGS) techniques (Zhao et al., 2014), became available for transcriptome studies and are independent of any annotated sequence from the organism of interest, with several advances from microarray technologies and several sequencing methods (Valdés et al., 2013; van Dijk et al., 2014). For non-model species lacking a reference genome, RNA-Seq is a powerful tool to reconstruct and quantify whole transcriptomes (Conesa et al., 2016; Hrdlickova et al., 2017). However, there is no consensus or optimal methodologies to use in transcriptome studies, thus creating a challenge to analyse such data. Therefore, different species, sequencing protocols, and parameter settings necessitate different approaches to achieve the most reliable results (Hölzer and Marz, 2019).

1.3.1 Transcriptome assembly and its challenges

Transcriptome assembly from RNA-Seq data is a useful method for detecting variations in gene expression and sequences, as well as finding novel transcripts or alternatively spliced transcripts between conditions, tissues and species (Wang et al., 2009; Ozsolak and Milos, 2011; Li et al., 2019a). There are two main approaches for transcriptome assemblies: i) reference-guided assembly, which takes advantage of an existing genome to which the RNA-Seq reads are first aligned and further used for the construction of individual transcripts; and ii) *de novo* assembly, which relies on overlaps between reads and attempts to assemble them into full transcripts, without aligning the reads to a reference genome (Ungaro et al., 2017; Kovaka et al., 2019). For non-model species that lack a high-quality reference genome, the *de novo* assembly approach is a powerful tool to aid transcriptome analysis. However, several problems are exacerbated due to the lack of reference sequences, such as the presence of chimaeras, structural and base errors, presence of partial transcripts and incomplete assemblies (Smith-Unna et al., 2016; Li et al., 2019a).

Although similarities are shared between genome and transcriptome assemblies, reconstructing a comprehensive and detailed *de novo* transcriptome from RNA-Seq data imposes many informatics challenges. Whereas the depth of DNA sequencing is similar across a genome, some transcripts might be highly expressed, while others present a low expression (Martin and Wang, 2011; Hölzer and Marz, 2019). Low expressed transcripts be might be considered as sequencing errors and removed from the assembly process by various tools (Haas and Zody, 2010), while too many reads lead to potential misassembly and increased runtimes, and therefore require *in silico* reduction of the number of reads (Haas et al., 2013; Conesa et al., 2016). Secondly, reads derived from one exon can be part of multiple paths in the assembly graph, thus making it difficult to resolve transcript variants unambiguously (Hölzer and Marz, 2019). Lastly, repetitive regions are also a major problem for the construction of transcripts, and coverage cannot be easily applied to discriminate contigs that correspond to repeats as it is done for genomic repeats (Novák et al., 2010; Lima et al., 2017; Hölzer and Marz, 2019).

Several tools have been developed for *de novo* transcriptome assembly in the last decades, for instance, Trans-ABySS (Robertson et al., 2010), IDBA-tran (Peng et al., 2010), Trinity (Grabherr et al., 2011), Oases (Schulz et al., 2012), SOAPdenovo-Trans (Xie et al., 2014), Bridger (Chang et al., 2015), BinPacker (Liu et al., 2016), Shannon (Kannan et al., 2016) and rnaSPAdes (Bushmanova et al., 2018). While most of these tools use *de Bruijn* graph strategies to construct, revise, and parse the assemblies, each method uses different thresholds and different assumptions to make decisions (Voshall and Moriyama, 2018). Several studies have comparatively assessed the performance of these tools for different plant species, e.g. *Camellia sinensis* (Zhao et al., 2011; Li et al., 2019a), Arabidopsis (Jain et al., 2013; Wang and Gribskov, 2017), peanut (Chopra et al., 2014), *Nicotiana benthamiana* (Nakasugi et al., 2014) and sugarcane (Hoang et al., 2017). Therefore, knowing the advantages and disadvantages of each tool, as well as adjusting the algorithms to obtain the best possible results, is an essential step in the direction of a successful and reliable *de novo* transcriptome assembly (Hölzer and Marz, 2019)

1.3.2 The power of Single-Molecule Sequencing for transcriptome assembly

While NGS technologies have revolutionized the field of molecular biology due to their massive improvements over Sanger sequencing, their limitations, such as the application

of short reads, make them poorly suited for some particular biological problems, including genome and transcriptome assemblies, gene variant detection and epigenetics studies (Rhoads and Au, 2015; An et al., 2018). As an alternative to overcome many of the NGS limitations, Third Generation Sequencing (TGS) presents distinguishing features, such as single-molecule sequencing (SMS), leading to the absence of PCR bias, sequencing in real-time and production of long reads (Bleidorn, 2016; van Dijk et al., 2018; Wang et al., 2019a). The first commercialized SMS technology was the Helicos Biosciences (Harris et al., 2008; Pushkarev et al., 2009), but it was not viable due to its high costs and production of short reads (van Dijk et al., 2018). The first successful one was released by Pacific Biosciences (PacBio), termed single-molecule real-time (SMRT) sequencing (Eid et al., 2009), and it was followed by the nanopore sequencing from Oxford Nanopore Technologies (ONT) (Jain et al., 2015).

The SMRT technology from PacBio produced initial average short read lengths (~1.5kb) with error rates around 13% (Quail et al., 2012); however, due to the development of sequencing chemistry and a release of a new sequencer, read length and throughput have substantially increased and led to an overall accuracy improvement (van Dijk et al., 2018). Its circular consensus sequencing mode (CCS), where molecules are sequenced several times, provide base-level resolution with 99% single-molecule read accuracy (Wenger et al., 2019). In contrast, nanopore reads are limited to the length of the DNA molecules to be sequenced rather than the chemistry used (van Dijk et al., 2018). Thus, unusual long reads can be obtained, up to ~1Mb (Jain et al., 2018), if sufficient template quality is available. Although a new nanopore sequencing method called 2D sequencing allows for the sequencing of both DNA strands, it is still no able to sequence the same strand multiple times as the SMRT and, therefore, presenting high error rates (Jain et al., 2017; van Dijk et al., 2018).

The application of PacBio long-read isoform sequencing (Iso-Seq) can provide long reads up to ~10 kb, thus retrieving most of the expressed transcripts as full-length sequences, alternative isoforms and duplicated genes (An et al., 2018). Besides, due to the high frequency of error of long PacBio reads, hybrid sequencing approaches, which integrate short reads with long reads during the transcriptome assembly, are valuable tools for plant transcriptome studies (Au et al., 2012; Koren et al., 2012; An et al., 2018). Such approaches have been used to unveil the complexity of the maize transcriptome (Wang et al., 2016), improve the wheat genome annotation (Dong et al., 2015), survey the transcriptomes of sorghum, sugarcane and coffee bean (Abdel-Ghany et al., 2016;

Cheng et al., 2017; Hoang et al., 2017), assess alternative splicing in cultivated and wild cotton (Wang et al., 2018; Feng et al., 2019), rhizome-associated alternative splicing in bamboo (Wang et al., 2017), study taxol biosynthesis in *Taxus cuspidate* (Kuang et al., 2019). Therefore, SMS is a useful approach to generate valuable resource to the plant research community, providing genome and transcriptome information at unprecedented resolutions and accuracy (An et al., 2018; van Dijk et al., 2018).

1.4 Mutation breeding

Genetic variation, a result of recombination of genetic materials and their continuous interaction with the environment, is the main force driving evolution and practical breeding (Oladosu et al., 2016). Plant breeding, the very basis of modern civilization, play a significant role in increasing the genetic variability for desired traits in various domesticated plant species (Sikora et al., 2011; Chaudhary et al., 2019). Classical breeding (i.e. recombinant breeding) remains the standard approach of the plant breeding industry (Acquaah, 2012). However, mutation breeding has been an extraordinary tool in plant breeding since the late 1920s, when Lewis John Stadler showed the applicability of X-ray in inducing genetic variation in barley and maize (Stadler, 1928a, 1928b).

Mutation breeding aims to induce genetic variation, by chemical, physical or biological agents (or mutagens), without major losses in viability (Roychowdhury and Tah, 2013; Oladosu et al., 2016). Chemical mutagens, such as N-methyl-N-nitrosourea (MNU), sodium azide, hydrogen fluoride (HF), methyl methanesulfonate (MMS), and ethyl methanesulfonate (EMS), often cause mild effects in plant tissues, and they have several advantages, including simplicity to handle, do not require specialised equipment, and can provide a very high mutation frequency (Sikora et al., 2011; Acquaah, 2012; Oladosu et al., 2016). EMS is the most widely used chemical mutagen in plants, as it produces point mutations with high frequency (specially G-C to A-T base pair transitions), rather than deletions and translocations (Greene et al., 2003; Till et al., 2007). Such mutations can occur in coding regions, in the form of silent, missense or nonsense mutation, as well as in non-coding regions, which can lead to changes in promoter sequences or other regulatory regions, affecting the regulation of transcription (Mba, 2013; Oladosu et al., 2016).

Physical mutagens are divided into two categories: i) ionizing atomic particle radiation, made up of subatomic particles (electrons, protons, neutrons), with fast neutron and ion beams being the most used for plant mutation induction (Li et al., 2001; Tanaka et al., 2010); and ii) ionizing electromagnetic radiations, which include cosmic rays, X-rays and gamma rays (Mba, 2013; Bado et al., 2015). Ionizing radiation induces DNA oxidative damage, such as double-strand breaks (DSBs), base substitutions, deletions and chromosomal alterations at a lower frequency, thus resulting in common loss of gene function (Morita et al., 2009; Kazama et al., 2011; Jo and Kim, 2019). DSBs can be repaired either via the error-free homologous recombination (HR) pathway or via the error-prone non-homologous end-joining (NHEJ) pathway, with the latter being the predominant DSB repair pathway in plants (Puchta, 2005; Bleuyard et al., 2006).

Mutation breeding has been used to generate varieties with different agronomical and quality traits for more than 200 different species and crops (FAO and IAEA, 2019). The development of varieties resistant to biotic and abiotic stresses is a major field of mutation breeding. EMS wheat mutants showed a broad-spectrum disease resistance (Campbell et al., 2012), while sodium azide mutagenesis generated the same in mutant rice (Wang et al., 2019b). Gamma-ray treated sesame varieties have shown field tolerance to *Phytophthora* (Pathirana, 1992). Drought and salinity tolerance were observed in gamma-ray irradiated wheat and rice (Sen et al., 2017; Abdelnour-Esquivel et al., 2020). In addition to that, improvement of crop quality and nutritional traits has also been achieved with mutation breeding, including increased quality in sweet cherry fruits (Kunter et al., 2012), increased levels of antioxidants in black gram (*Vigna mungo* L. Hepper) (Yasmin et al., 2019), high-amylose cassava mutants (Ceballos et al., 2008) and high oleate levels in peanut (Mondal et al., 2011).

1.4.1 Mutation breeding in ornamentals

Mutation breeding is an established and successful method to generate variation in ornamental plants, since most economically important traits, e.g. flower characteristics or growth habit, are easily monitored after the mutagenic treatment (Schum and Preil, 1998; Ibrahim et al., 2018). Induced mutations in ornamentals focus on economic aesthetic values, such as flower and leaf characters (colour, size, morphology, fragrance), growth habit (compact, climbing, branching), physiological traits (photoperiodic response, early flowering, free-flowering, flower keeping quality), and

biotic and abiotic stress responses (Schum and Preil, 1998). Furthermore, most ornamental species are vegetatively propagated and highly heterozygous; this allows for the detection, selection and maintenance of mutants in early generations of the breeding process, without drastically altering other desired characteristics of a genotype (Broertjes and Van Harten, 1988; Schum and Preil, 1998; Van Harten, 2002).

The use of mutation breeding in ornamental plants has started in the 1930s, where W.E. de Mol first irradiated tulip bulbs (cv. Fantasy), which had as an outcome the first commercial mutant released with altered flower colour (cv. Faraday) at the end of the 1940s (Broertjes and Van Harten, 1988). Over the years, the applicability of induced mutagenesis has been exploited in several ornamental species (Datta, 2020), with over 700 mutant varieties for more than 80 different species (FAO and IAEA, 2019). The largest number of mutated varieties has been obtained in the genera *Chrysanthemum*, *Rosa*, *Dahlia*, *Alstroemeria* and *Streptocarpus* (Table 2). EMS, X-ray, gamma-ray, and ion-beam irradiation were used to produce Chrysanthemum cultivars with novel traits, especially new flower colours, shapes and sizes (Broertjes, 1966; Rocha Latado et al., 2004; Silva et al., 2013; Soliman et al., 2014; Okamura et al., 2015). Many rose varieties have been developed through both physical and chemical mutagens (reviewed by (Datta, 2018). Ion-beam irradiation was applied to obtain carnation mutants with different flower colour and shapes (Okamura et al., 2003, 2013).

Table 2. Top 10 genera with the largest mutant varieties registered in the joint Food and Agriculture Organization of the United Nations/International Atomic Energy Agency database (FAO and IAEA, 2019).

Species	Mutated varieties obtained	Physical mutagen	Chemical mutagen
Chrysanthemum	285	271	1
Rosa	67	57	4
Dahlia	35	35	0
Alstroemeria	35	35	0
Streptocarpus	30	27	1
Dianthus	28	21	3
Begonia	25	25	0
Portulaca	17	17	0
Bougainvillea	16	11	1
Rhododendron	15	15	0
Others	174	147	16

1.4.2 Microsatellites as a source of variation

Microsatellites, Single Sequence Repeats (SSRs) or Short Tandem Repeats (STRs), are among the most variable and highly dynamic types of repetitive sequence in the genome (Ellegren, 2004; Verstrepen et al., 2005). They are defined as short tandemly repeated DNA sequences that involve a repetitive unit of 1-9 bp and lengths of up to 100 nucleotides (nt) (Fan and Chu, 2007; Gemayel et al., 2010). SSR regions are extremely unstable, where mutations occur in the form of an addition or deletion of repeat units at much higher rates than in other parts of the genome (10 to 100,000 times higher) (Schlötterer, 2000; Gemayel et al., 2010). The majority of SSR mutations occur due to replication slippage, which happens during DNA replication, leading to mispairing (by one or more repeat units) between the nascent and template strands (Fan and Chu, 2007; Gemayel et al., 2010) (Figure 3). However, mutations in SSR locations can also arise from DNA double-strand break (DSB) repair (Pâques et al., 1998; Richard and Pâques, 2000; Verstrepen et al., 2005).

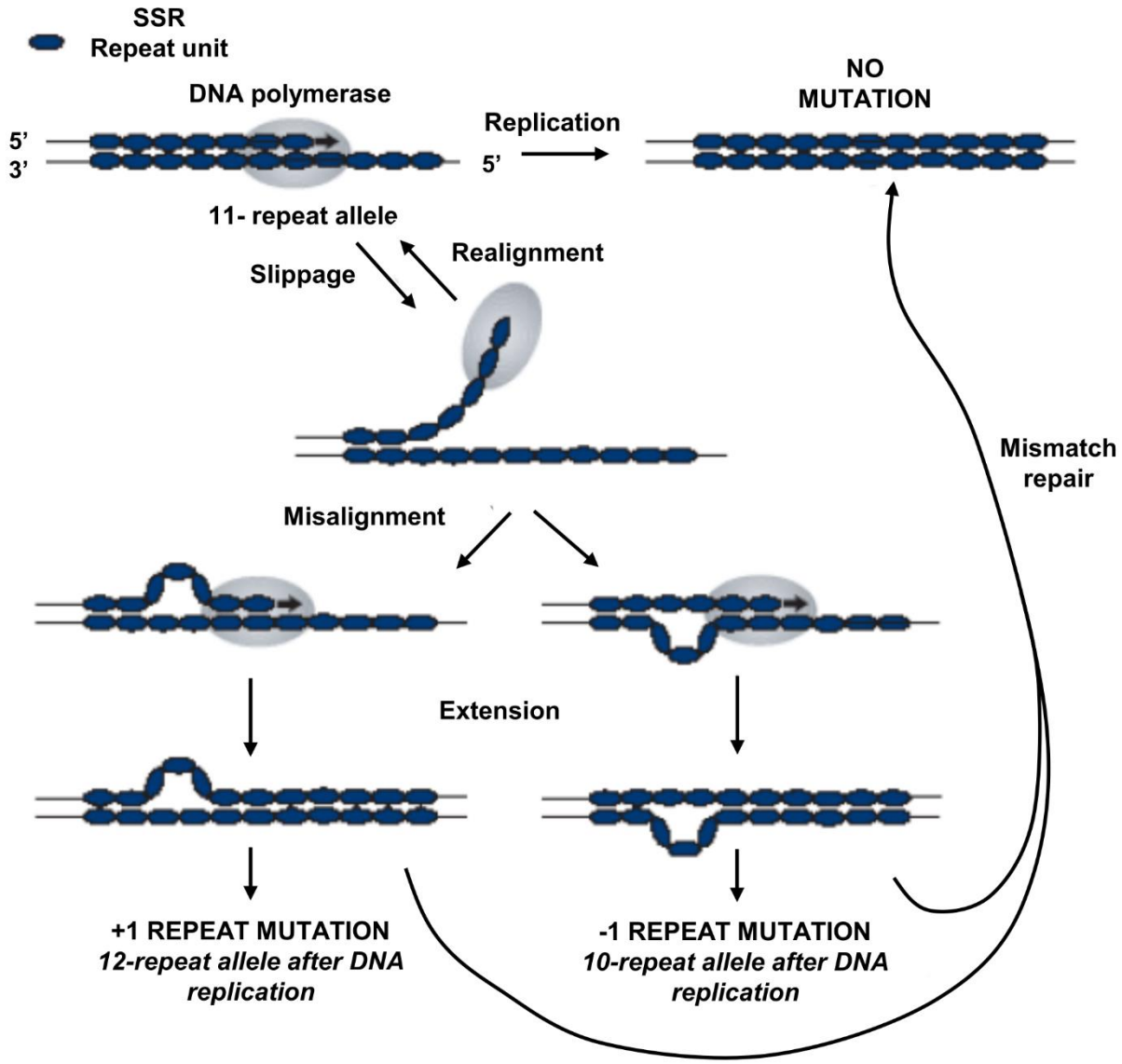


Figure 3. Schematic illustration of replication slippage during DNA replication. Misalignment of nascent and template strands can either lead to an addition or deletion of one or more repeat units in the newly replicated strand (adapted from Fan and Chu (2003)).

Ionizing radiation (e.g. X-rays and γ-rays) leads to DSBs, due to DNA oxidative damage, which can be subsequently repaired either via the error-free homologous recombination (HR) pathway or via the error-prone non-homologous end-joining (NHEJ) pathway, with the latter being the predominant DSB repair pathway in plants (Puchta, 2005; Bleuyard et al., 2006; Roldán-Arjona and Ariza, 2009). Nonetheless, HR has been shown to occur in low frequencies, where information from both allelic and ectopic sites could be used for DSB repairs (Puchta, 1999; Gisler et al., 2002; Orel et al., 2003). If homologous regions close to the DSB are present, HR can occur in higher frequency (Siebert and

Puchta, 2002). The classical NHEJ (cNHEJ) involves the ligation of the DSB junctions without much sequence loss, which is due to the fact that double-stranded ends are protected from degradation by binding of a KU-heterodimer (KU70/KU80) before ligation takes place (Knoll et al., 2014).

On the other hand, the alternative NHEJ (aNHEJ) can take place when, at the site of the DSB, single strands containing microhomologies are present and align to one another (Knoll et al., 2014). This type of aNHEJ, also known as microhomology-mediated end joining (MMEJ), is associated with deletions flanking the original DSB site (Sfeir and Symington, 2015). Nonetheless, deletions can also occur via HR-mediated DSB repair. Single-strand annealing (SSA) depends on homologous repetitive sequences that are arranged nearby, and they showed to be as efficient as NHEJ in DSB repair in plants (Siebert and Puchta, 2002; Orel et al., 2003; Vu et al., 2017). Studies have shown that SSR instability increases with plant development in Arabidopsis and that NHEJ repair is more efficient in such cases (Golubov et al., 2010). Moreover, abiotic stress has also been shown to induce SSR instability, which might be linked with the frequency of errors made during NHEJ repair of DSBs (Yao and Kovalchuk, 2011).

Despite their importance in many fields of molecular biology, reliable SSR genotyping from Next Generation Sequencing (NGS), especially for small insertions and deletions (*indels*) analysis, is still challenging (Albers et al., 2011; Treangen and Salzberg, 2011; Wang et al., 2011). Several NGS methods have been used for genome-wide and targeted SSR genotyping (Vartia et al., 2016; De Barba et al., 2017; Li et al., 2017, 2019b; Yang et al., 2019). However, there is an increased sequencing error in such regions due to imprecise base calling of NGS platforms and polymerase slippage during PCR amplification (Albers et al., 2011). Studies using plasmids containing artificially inserted tandem repeat sequences have shown the appearance of stutters due to slippage, which is strongly correlated to the number of repeat units and repeat unit length (Shinde et al., 2003; Fungtammasan et al., 2015). Nonetheless, the information obtained by NGS offers a better and more powerful resolution than the traditional ways of SSR genotyping (Šarhanová et al., 2018).

1.5 Objectives

Anthocyanins are well-known plant pigments responsible for red pigmentation in flowers and fruits. Their metabolic pathways, regarding enzymes and the encoding genes, have been extensively studied over the past years. In ornamental plants, the appearance of acyanic (uncoloured) mutants is common, which occurs due to the lack of enzyme expression in the biosynthetic pathways. However, in acyanic poinsettia (*Euphorbia pulcherrima*), gene expression and enzyme activities for most known factors investigated so far are unaltered. This phenomenon is commonly called the *"white paradox"*.

Furthermore, the colour range in poinsettia varieties is usually obtained by mutation breeding (i.e. radiation), thus generating a spectrum of bract colours, such as pink, marble, orange and white/creamy. The white varieties are often obtained through radiation mutagenesis of the red varieties, followed by shoot development and trait selection. Therefore, red and white poinsettias from the same variety are referred to as 'pairs', due to their highly similar genetic background. However, not all red varieties can produce white sports through radiation. Thus, red poinsettia varieties are distinguished into 'heterozygous' and 'homozygous' for the colouration locus according to their ability to generate white sports (heterozygous produce white sports and homozygous do not).

The following thesis objectives were determined which were analysed in the three manuscripts presented in this thesis:

1. to identify the genetic factors responsible for the white phenotype in poinsettia by assembling and functionally annotating a hybrid *de novo* bract transcriptome for the species, as well as to characterize the main differences between red- and white-bracted poinsettia varieties during bract development, with a focus in the flavonoid and adjacent pathways that are involved in pigment accumulation in plant tissues (manuscript 1).
2. to functionally characterize an anthocyanin related *GST* from poinsettia as the most likely target of the radiation-induced mutation of red poinsettias into white bract sports (manuscript 2).
3. to identify a molecular marker and develop an approach to detect a low-frequency deletion mutation in the poinsettia *GST*, and thus improve selection in poinsettia breeding programs (manuscript 3).

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

Vinicius Vilperte[1,2†], Calin Rares Lucaciu[3†], Heidi Halbwirth[4], Robert Boehm[2], Thomas Rattei[3] and Thomas Debener[1]

[1]Institute for Plant Genetics, Leibniz University Hannover, Hannover, Germany

[2]Klemm + Sohn GmbH & Co., Stuttgart, Germany

[3]Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria

[4]Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, Vienna, Austria

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

Type of authorship:	First author
Type of article:	Research article
Contribution to the article:	Planned and performed most of the experiments, analysed the data, prepared all tables and figures, and wrote most of the manuscript.
Contribution of the other authors:	Calin Rares Lucaciu performed most of the bioinformatic analysis and helped with the writing of the manuscript. Heidi Halbwirth, Robert Boehm, and Thomas Rattei helped in conceptualizing the project and reviewing the manuscript. Thomas Debener contributed to the planning of the experimental design, data analysis, and writing and reviewing of the manuscript.
Supplements:	Supplementary material mentioned in the manuscript is provided at the end of the chapter or as an electronic appendix.
Journal:	BMC Genomics
Date of publication:	27.11.2019
Impact factor:	4.093
DOI:	https://doi.org/10.1186/s12864-019-6247-3

Vilperte *et al. BMC Genomics* (2019) 20:900
https://doi.org/10.1186/s12864-019-6247-3

BMC Genomics

RESEARCH ARTICLE Open Access

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

Vinicius Vilperte[1,2†], Calin Rares Lucaciu[3†], Heidi Halbwirth[4], Robert Boehm[2], 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 Lucaciu contributed equally to this work.
[3]Department of Microbiology and Ecosystem Science, University of Vienna, 1090 Vienna, Austria
[1]Institute 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,

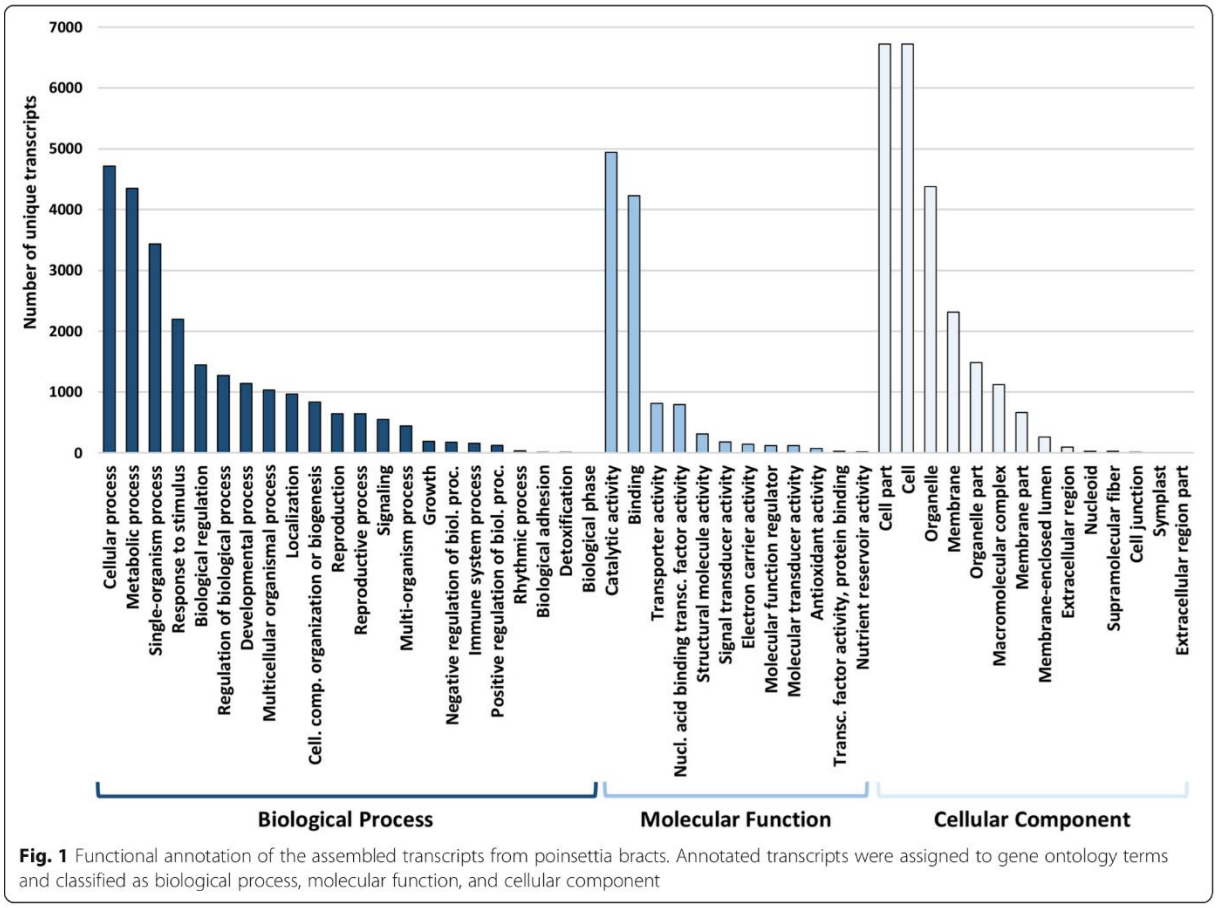


Fig. 1 Functional annotation of the assembled transcripts from poinsettia bracts. Annotated transcripts were assigned to gene ontology terms and classified as biological process, molecular function, and cellular component

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

	A. thaliana orthologous	Gene	Enzyme name	# of transcripts identified
Structural genes	AT2G37040.1	*PAL1*	Phenylalanine ammonia-lyase 1	4
	AT3G53260.1	*PAL2*	Phenylalanine ammonia-lyase 2	5
	AT3G10340.1	*PAL4*	Phenylalanine ammonia-lyase 4	1
	AT2G30490.1	*C4H*	Trans-cinnamate-4-hydroxylase	8
	AT1G51680.1	*4CL1*	4-coumarate: CoA ligase 1	3
	AT3G21240.1	*4CL2*	4-coumarate: CoA ligase 2	5
	AT1G65060.1	*4CL3*	4-coumarate: CoA ligase 3	4
	AT3G21230.1	*4CL5*	4-coumarate: CoA ligase 5	2
	AT5G13930.1	*CHS*	Chalcone synthase	4
	AT5G48930.1	*HCT*	Hydroxycinnamoyl-CoA shikimate transferase	12
	AT5G05270.1	*CHI3*	Chalcone-flavonone isomerase 3	2
	AT3G55120.1	*CHI1*	Chalcone-flavonone isomerase 1	4
	AT3G51240.1	*F3H, FHT*	Flavanone 3-hydroxylase	1
	AT5G07990.1	*F3'H*	Flavonoid 3′-hydroxylase	10
	AT5G08640.1	*FLS1*	Flavonol synthase 1	5
	AT5G42800.1	*DFR*	Dihydroflavonol 4-reductase	11
	AT4G22880.1	*LDOX, ANS*	Leucoanthocyanidin dioxygenase, Anthocyanin synthase	3
	AT5G17050.1	*UGT78D2*	UDP-glucosyl transferase 78D2	10
	AT3G29590.1	*5MAT*	Anthocyanidin 5-*O*-glucoside-6″-*O*-malonyltransferase	2
	AT1G61720.1	*ANR*	Anthocyanidin reductase	10
	AT5G54160.1	*OMT1*	Flavone 3′-*O*-methyltransferase 1	15
	AT3G59030.1	*TT12*	TRANSPARENT TESTA 12	5
	AT3G03190.1	*GSTF11*	Glutathione S-transferase	4
Transcription factors	AT1G63650.1	*EGL1*	Transcription factor EGL1	7
	AT1G66370.1	*MYB113*	MYB domain protein 113	2
	AT2G47460.1	*MYB12*	MYB domain protein 12	4
	AT5G35550.1	*TT2*	Transcription factor TT2 (MYB123)	6
	AT4G09820.1	*TT8*	Transcription factor TT8 (bHLH42)	4
	AT5G24520.1	*TTG1*	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		
E. pulcherrima - Bract	41.3%	36.1%	9.7%	12.9%
E. pekinensis - Leaf	31.3%	50.1%	8.9%	9.7%
E. pekinensis - Root	32.0%	46.5%	10.2%	11.3%
E. pekinensis - 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)

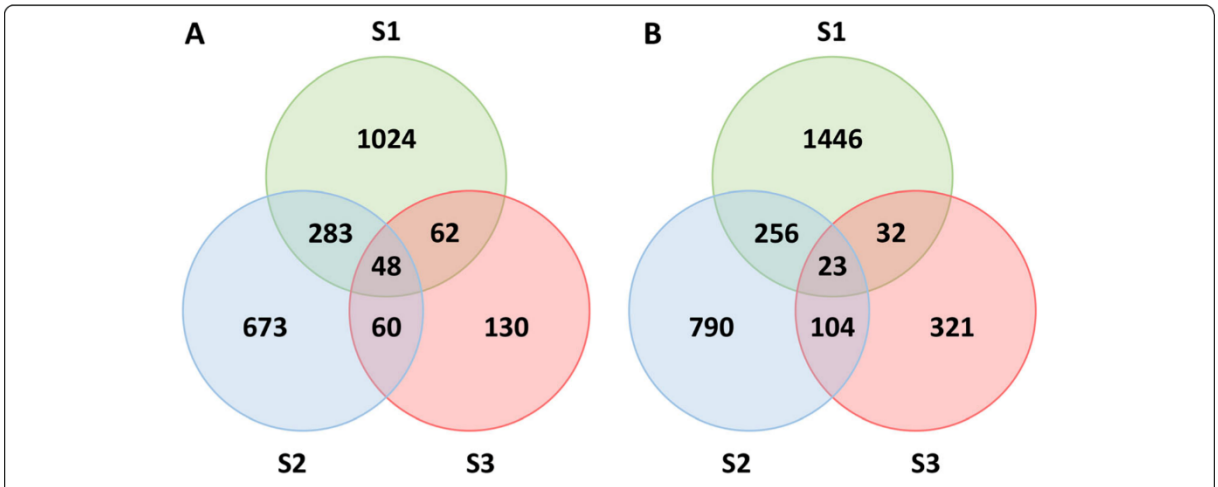


Fig. 2 Venn diagram of the differentially regulated transcripts for the different bract developmental stages of poinsettia. **a** Transcripts with a lower expression in white bracts; **b** Transcripts with a higher expression in white bracts. S1, S2 and S3 = Stages 1, 2 and 3, respectively

and jasmonic acid biosynthetic/metabolic processes (GO:0009695/GO:0009694). Moreover, several molecular functions related to transferase and glucosyltransferase/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

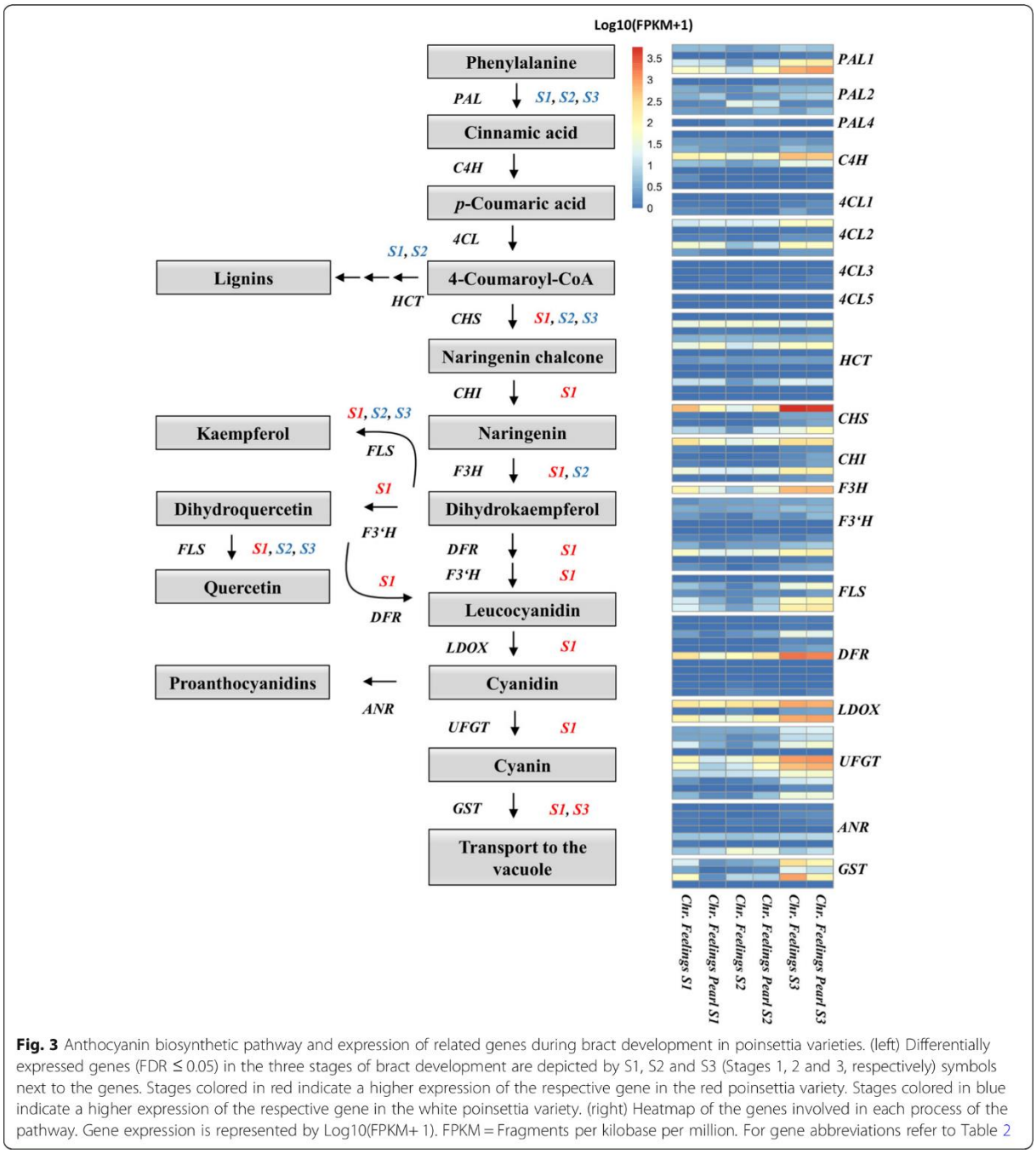


Fig. 3 Anthocyanin biosynthetic pathway and expression of related genes during bract development in poinsettia varieties. (left) Differentially expressed genes (FDR ≤ 0.05) in the three stages of bract development are depicted by S1, S2 and S3 (Stages 1, 2 and 3, respectively) symbols next to the genes. 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*, *UFGT*, *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 Pffafl 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 caerulescens* - 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.

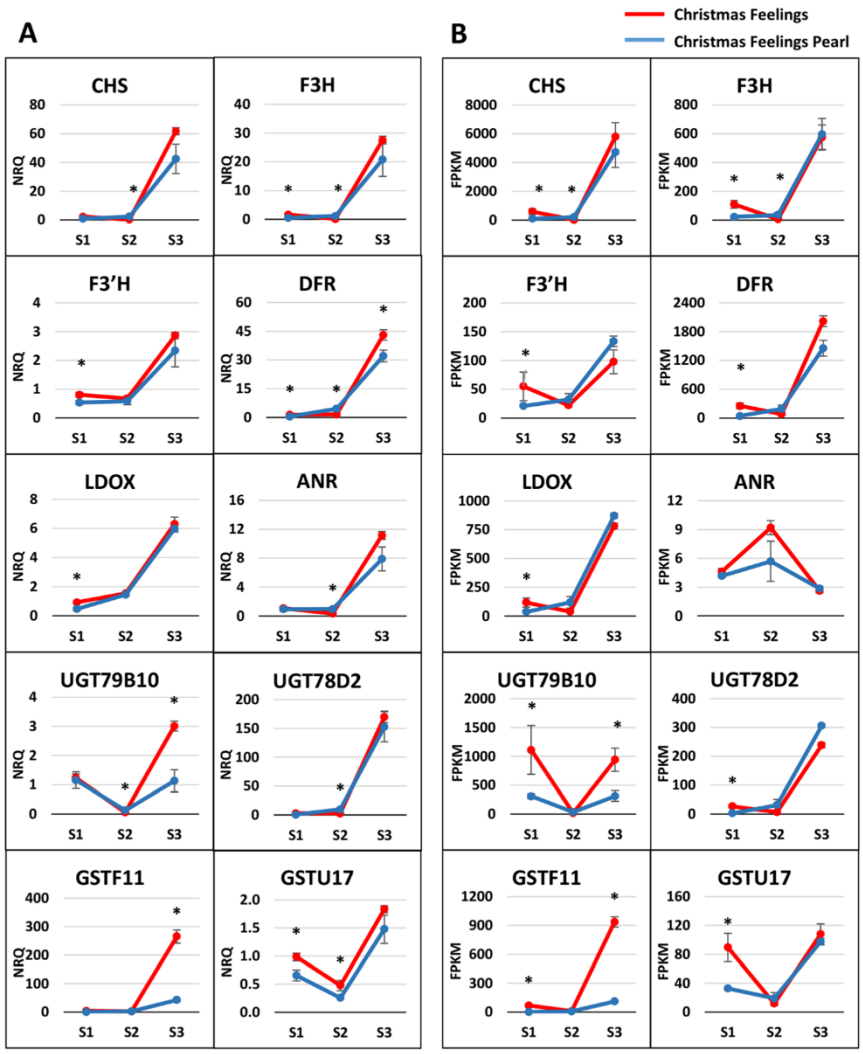


Fig. 4 Expression profiles of anthocyanin-related genes for three developmental stages of poinsettia bracts. **a** RT-qPCR expression profiles of 10 anthocyanin related genes for the varieties Christmas Feelings and Christmas Feelings Pearl in three stages of bract development. **b** RNA-Seq expression 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 *single-end*	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 *paired-end*	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 (Cellscript, 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

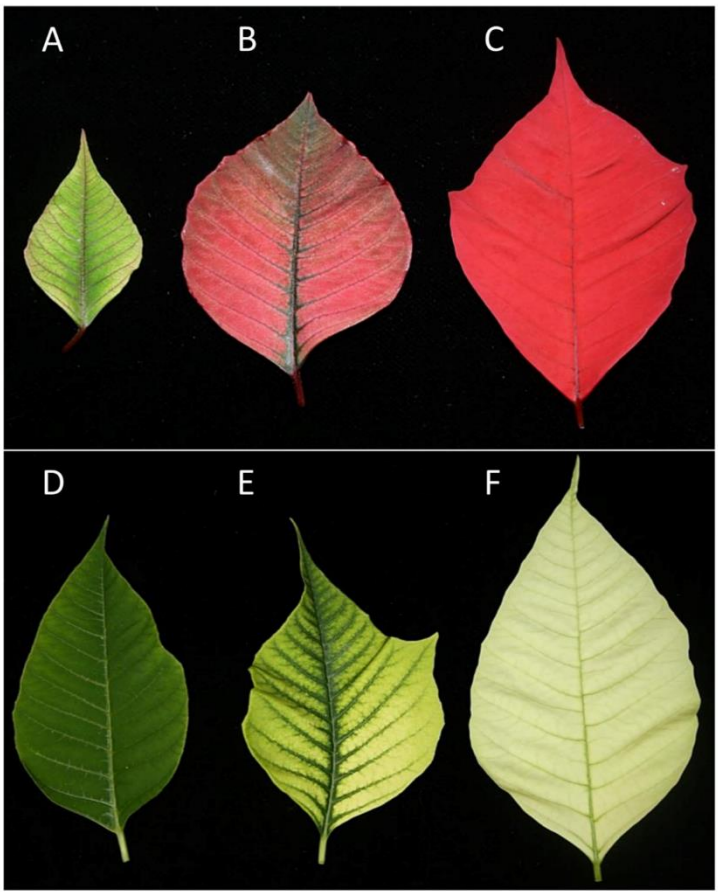


Fig. 5 Bracts of red and white poinsettia varieties for three color developmental stages. **a-c** Bracts from the Christmas Feelings variety for stages 1, 2 and 3, respectively; **d-f** Bracts from the Christmas Feelings Pearl variety for stages 1, 2 and 3, respectively

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) log2FC ≥ 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. VV, CL, HH, RB, TR, and TD designed the experiments. VV and CL conducted the bioinformatics analyses. VV performed the RT-qPCRs. VV 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.

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
[1]Institute of Plant Genetics, Leibniz Universität Hannover, 30419 Hannover, Germany. [2]Klemm + Sohn GmbH & Co., 70379 Stuttgart, KG, Germany. [3]Department of Microbiology and Ecosystem Science, University of Vienna, 1090 Vienna, Austria. [4]Institute 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

1. 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 SlMYB12-regulated transcriptional network. PLoS Genet. 2009;5:e1000777. https://doi.org/10.1371/journal.pgen.1000777.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.1365-313X.2004.02138.x.
11. 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.
12. 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.
13. Blande D, Halimaa P, Tervahauta AI, Aarts MGM, Kärenlampi SO. De novo transcriptome assemblies of four accessions of the metal hyperaccumulator plant Noccaea caerulescens. Sci Data. 2017;4:160131. https://doi.org/10.1038/sdata.2016.131.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Chalker-Scott L. Environmental significance of Anthocyanins in plant stress responses. Photchem Photbio. 1999;70:1. https://doi.org/10.1111/j.1751-1097.1999.tb01944.x.
20. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu 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.
21. Czemmel S, Stracke R, Weisshaar B, Cordon N, Harris NN, Walker AR, et al. The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. Plant Physiol. 2009;151:1513–30. https://doi.org/10.1104/pp.109.142059.
22. 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.
23. 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.1365-313X.2006.02964.x.
24. 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

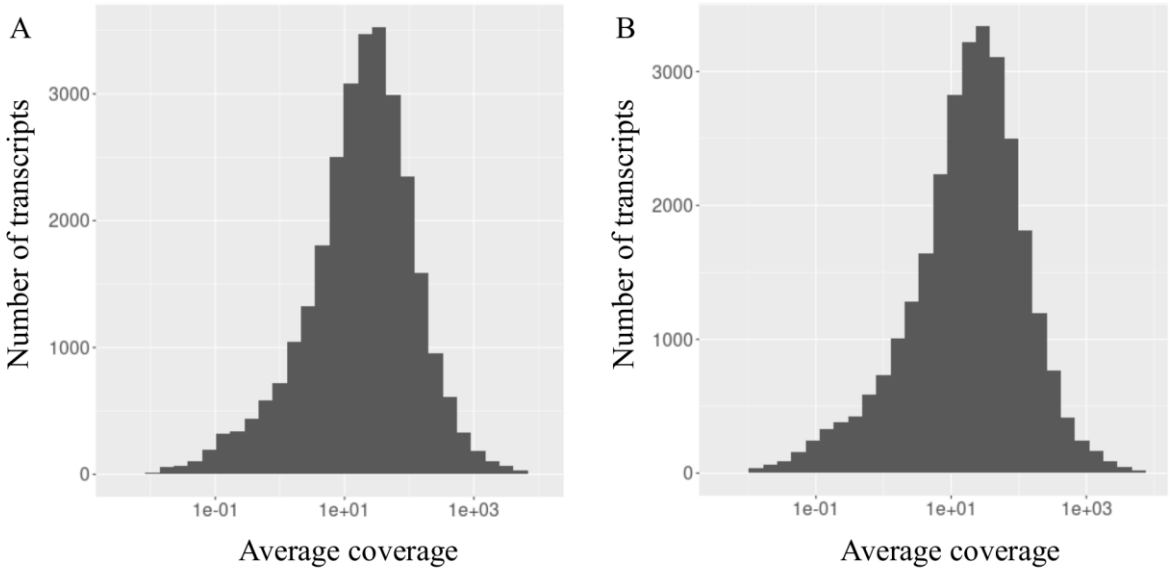
(Prunus salicina Lindl.) during fruit ripening using RNA-Seq. Front Plant Sci. 2016;7:1338. https://doi.org/10.3389/fpls.2016.01338.
25. Fu S, Ma Y, Yao H, Xu Z, Chen S, Song J, et al. IDP-denovo: de novo transcriptome assembly and isoform annotation by hybrid sequencing. Bioinformatics. 2018;34:2168–76. 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.1365-313X.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 Klotzsch) 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.
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.1365-313X.2003.01943.x.
39. Kobayashi S, Ishimaru M, Ding CK, Yakushiji H, Goto N. Comparison of UDP-glucose:flavonoid 3-O-glucosyltransferase (UFGT) 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.
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, Marcelis 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. Mehrtens 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.1365-313X.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 G, 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/genes9010038.
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, Veberic 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, Eguiarte 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, Seppey 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. Yahyaa 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/gkw1119.
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 chinesis 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

2.1 Supporting Information

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. (Available as an electronic appendix).

Additional File 3. Distribution of per cent length coverage for the top matching database entries from the poinsettia bract transcripts. Hit_pct_cov_bin = hit percent coverage in the bin, Count_in_bin = Transcripts counts in the bin.

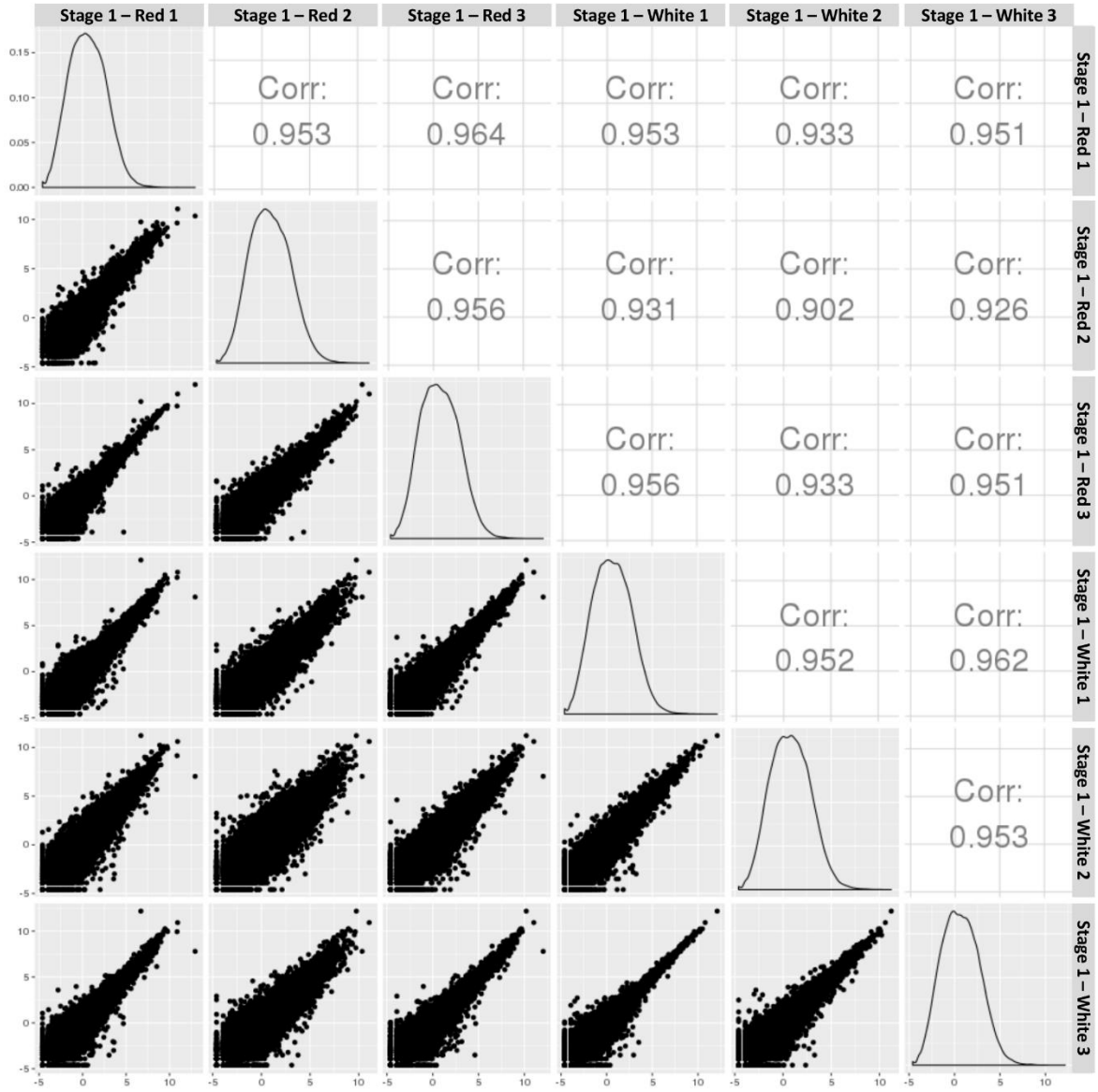
Hit_pct_cov_bin	count_in_bin	>bin_below
100	6,105	6,105
90	1,764	7,869
80	1,187	9,056
70	1,118	10,174
60	1,010	11,184
50	958	12,142
40	851	12,993
30	819	13,812
20	695	14,507
10	116	14,623

Additional File 4. BUSCO results for the *E. pulcherrima* and *E. pekinensis* transcriptomes. (Available as an electronic appendix).

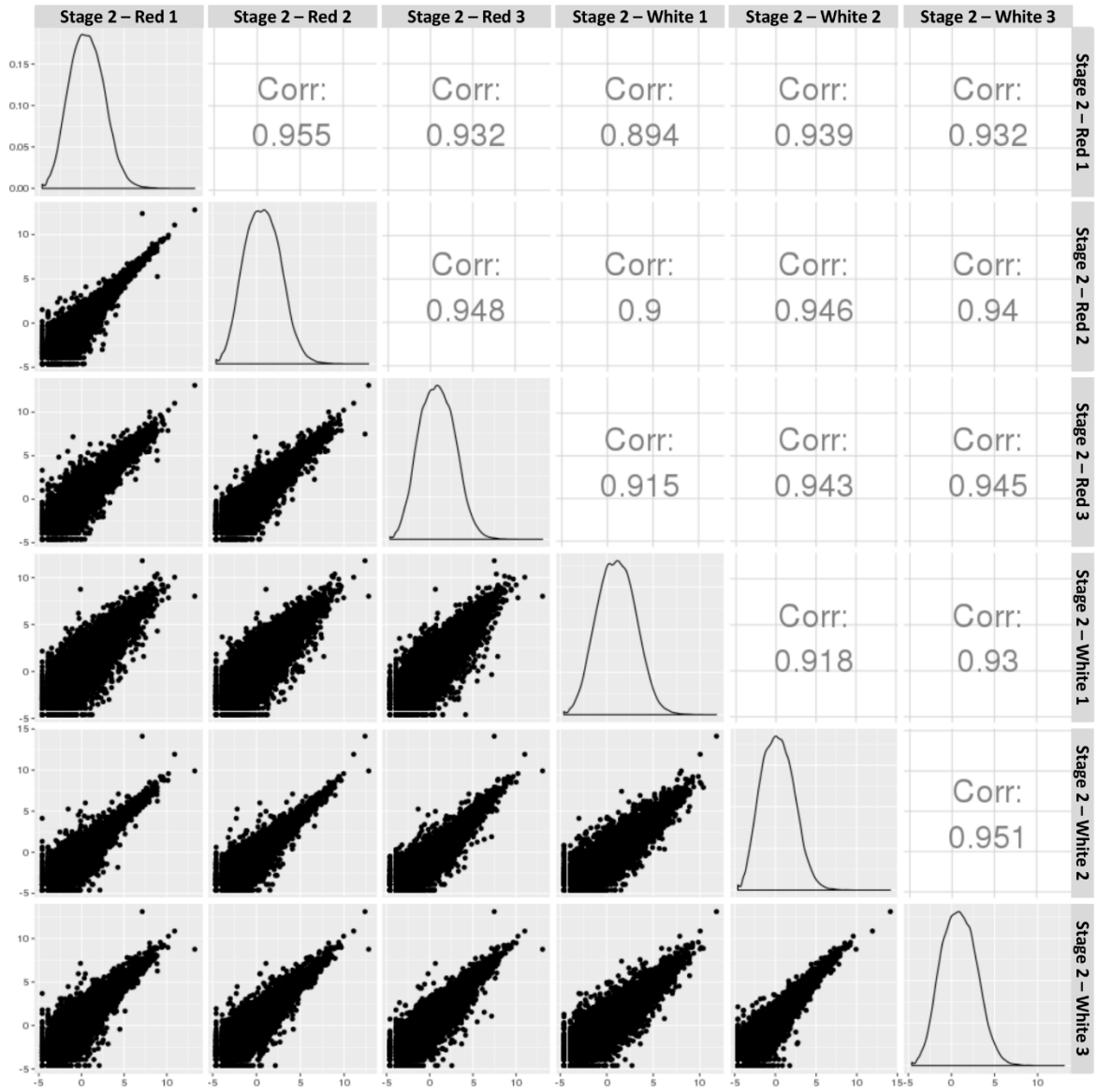
Additional File 5. Summary of the processed libraries for the Christmas Feelings and Christmas Feelings Pearl poinsettia varieties in three bract developmental stages. CF = Christmas Feelings, CFP = Christmas Feelings Pearl, R1 = Biological replicate 1, R2 = Biological replicate 2, R3 = Biological replicate 3.

Dev. Stage	Sample	Total number of reads	Remained reads after rRNA removal	Remained reads after quality trimming (QV≥20)	% of reads mapped to *E. pulcherrima* transcriptome
Stage 1	CF – R1	53,828,766	53,509,948	49,373,006	93.63
	CF – R2	55,591,532	55,260,594	51,259,165	88.79
	CF – R3	56,165,316	55,832,786	51,738,600	92.46
	CFP – R1	60,740,120	60,390,212	55,397,479	93.21
	CFP – R2	43,629,624	43,341,119	39,888,742	93.32
	CFP – R3	41,874,023	41,607,527	38,325,196	93.49
Stage 2	CF – R1	49,989,135	49,661,431	45,853,495	95.76
	CF – R2	53,357,892	52,994,339	49,099,190	94.53
	CF – R3	40,968,381	40,755,380	37,605,885	94.29
	CFP – R1	62,975,737	62,490,818	57,365,997	93.07
	CFP – R2	36,162,144	36,019,519	32,783,237	96.05
	CFP – R3	54,290,323	54,021,108	49,652,847	94.14
Stage 3	CF – R1	51,334,105	50,887,244	47,172,416	91.21
	CF – R2	55,486,852	55,147,781	51,264,695	91.71
	CF – R3	45,812,671	45,485,012	42,329,632	90.82
	CFP – R1	52,352,696	51,983,925	47,905,803	91.25
	CFP – R2	54,769,449	54,392,160	49,901,799	92.28
	CFP – R3	58,231,267	57,864,889	53,265,475	92.05

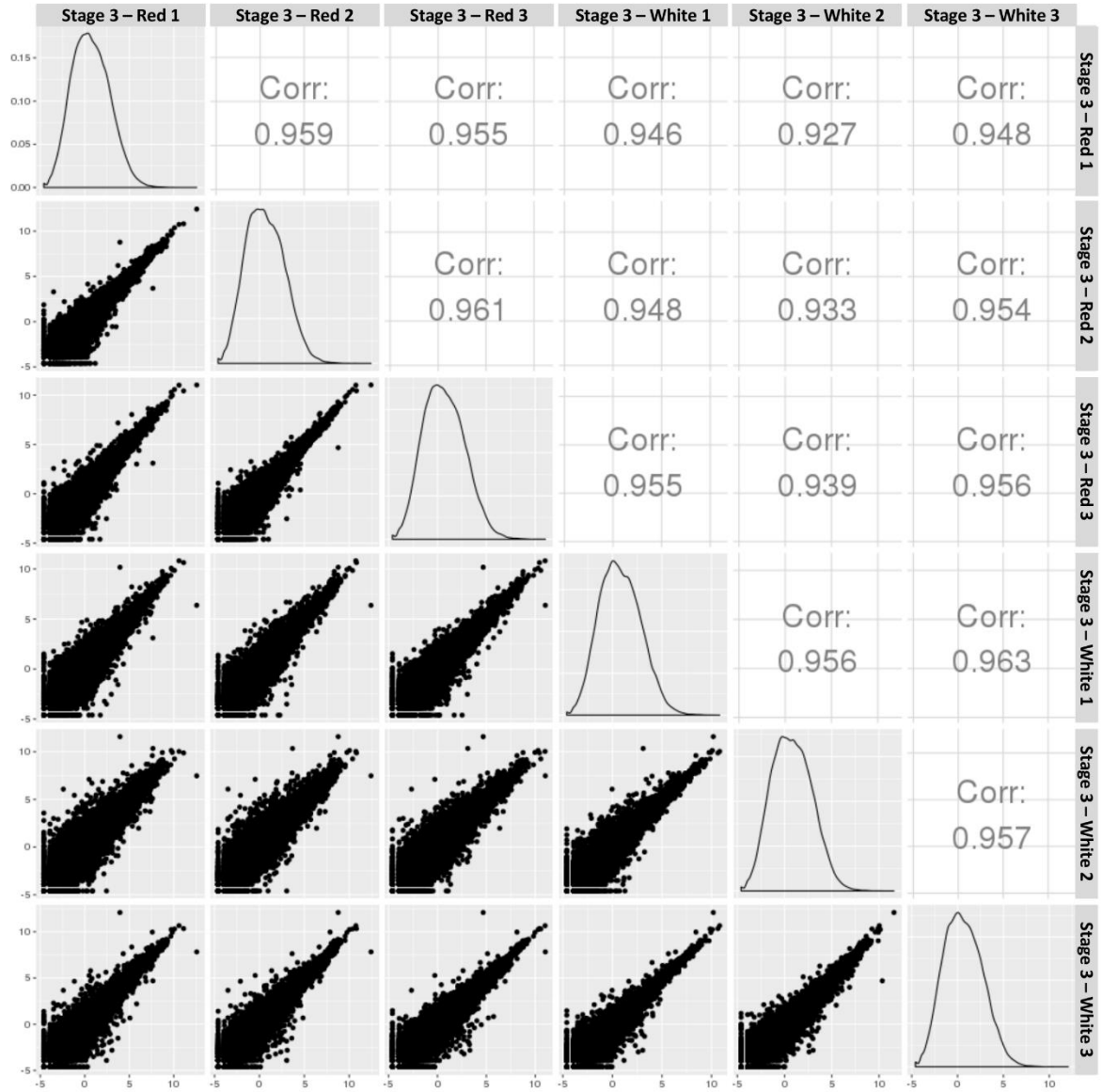
Additional File 6.1. Pearson correlation for the biological replicates of Christmas Feelings and Christmas Feelings Pearl for the first stage of bract development. Red 1, 2 and 3 = Biological replicates 1, 2 and 3 from the variety Christmas Feelings. White 1, 2 and 3 = Biological replicates 1, 2 and 3 from the variety Christmas Feelings Pearl.



Additional File 6.2. Pearson correlation for the biological replicates of Christmas Feelings and Christmas Feelings Pearl for the second stage of bract development. Red 1, 2 and 3 = Biological replicates 1, 2 and 3 from the variety Christmas Feelings. White 1, 2 and 3 = Biological replicates 1, 2 and 3 from the variety Christmas Feelings Pearl.



Additional File 6.3. Pearson correlation for the biological replicates of Christmas Feelings and Christmas Feelings Pearl for the third stage of bract development. Red 1, 2 and 3 = Biological replicates 1, 2 and 3 from the variety Christmas Feelings. White 1, 2 and 3 = Biological replicates 1, 2 and 3 from the variety Christmas Feelings Pearl.



Additional File 7. Differentially expressed transcripts for the comparisons between bract developmental stages in poinsettia. (Available as an electronic appendix).

Additional File 8. List of enriched GO terms for the comparisons between bract developmental stages in poinsettia. (Available as an electronic appendix).

Additional File 9. List of transcripts present in the differentially enriched pathways related to photosynthesis and phenylpropanoid metabolic processes. (Available as an electronic appendix).

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. (Available as an electronic appendix).

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. (Available as an electronic appendix).

Additional File 12. Primer sequences for each of the genes analysed by RT-qPCR in poinsettia bracts. *CHS = Chalcone synthase*, *F3H = Flavanone 3-hydroxylase*, *F3'H = Flavonoid 3'-hydroxylase*, *DFR = Dihydroflavonol 4-reductase*, *ANR = Anthocyanidin reductase*, *LDOX/ANS = Leucoanthocyanidin dioxygenase*, *UGT79B10 = UDP-glycosyltransferase 79B10*, *UGT78D2 = UDP-glycosyltransferase 78D2*, *GSTF11 = Glutathione S-transferase F11*, *GSTU17 = Glutathione S-transferase U17*, *EF1B = Translation elongation factor 1 beta*, *EF1A = Translation elongation factor 1 alpha*.

Transcript ID	TAIR10	Gene	Primers
TRINITY_DN44702_ c2_g2_i1	AT5G13930.1	*CHS*	F: CCACCGTCTCAACAGTCAAA R: GGAAGAAGTCTGCCGAAGATG
TRINITY_DN40768_ c0_g1_i1	AT3G51240.1	*F3H*	F: CCTTAGCCAGGCGAGTAATTT R: AAGAGGGAGGAGATTTGTAGGA
TRINITY_DN41584_ c0_g1_i1	AT5G07990.1	*F3'H*	F: CAGTACTGGAGTAATAAGCAACGA R: CCTAGGTTGTCGTCTCTAGTCTAT
TRINITY_DN50023_ c7_g1_i1	AT5G42800.1	*DFR*	F: GTTGAAACATGCCTAGCAAAGG R: CAGAGACAATTAGCCACAGGAA
TRINITY_DN44545_ c1_g6_i1	AT1G61720.1	*ANR*	F: TCTATGGTAGGTTAATATTTGAGGAGAG R: CAAACCACCAACATGAGGAATATC
TRINITY_DN52819_ c1_g1_i1	AT4G22880.1	*LDOX/ANS*	F: GAGCCGCCCAAGGATAAA R: AAAGTACGAGGAGGAAAGCG
TRINITY_DN37144_ c0_g2_i1	AT1G64910.1	*UGT79B10*	F: CGTACGACAACAAAGCTCAAAC R: TGGTGGATGTTGAGAAAGAGAC
TRINITY_DN48664_ c2_g1_i1	AT5G17050.1	*UGT78D2*	F: AACGGGTCAGCATCCCCGA R: GATGGAAATAATTTCTATGGCTTG
TRINITY_DN47136_ c7_g1_i2	AT3G03190.1	*GSTF11*	F: CTTGTTCATGAGCTTGGGTTAG R: CGGTCGGTCCGATATATCAAT
TRINITY_DN49420_ c1_g1_i1	AT1G10370.1	*GSTU17*	F: GGCTTAGAGTTACCGAGGTTTC R: TCAGGCAAAACATCCTTAACAG
TRINITY_DN42591_ c5_g2_i1	AT5G19510.1	*EF1B*	F: GGAGATGAGACTGAGGAGGATAA R: CCCACGGCTTTACATCCAATA
Jatropha curcas (Zhang et al., 2013).		*EF1A*	F: AAGATGATTCCCACCAAGCCCA R: CACAGCAAAACGACCCAGAGGA

3 A highly mutable GST is essential for bract colouration in *Euphorbia pulcherrima* Willd. Ex Klotsch

Vinicius Vilperte [1], Robert Boehm [2] and Thomas Debener [1]

[1] Institute for Plant Genetics, Leibniz University Hannover, Hannover, Germany

[2] Klemm + Sohn GmbH & Co., KG, 70379 Stuttgart, Germany

Type of authorship:	First author
Type of article:	Research article
Contribution to the article:	Planned and performed most of the experiments, analysed the data, prepared all tables and figures, and wrote most of the manuscript.
Contribution of the other authors:	Robert Boehm and Thomas Debener helped in conceptualizing the project and reviewing the manuscript. Thomas Debener contributed to the planning of the experimental design, data analysis, and writing and reviewing of the manuscript.
Supplements:	Supplementary material mentioned in the manuscript is provided at the end of the chapter or as an electronic appendix.

Journal:	BMC Genomics
Date of publication:	23.03.2021
Impact factor:	4.093
DOI (Research Square):	https://doi.org/10.1186/s12864-021-07527-z

Vilperte *et al. BMC Genomics* (2021) 22:208
https://doi.org/10.1186/s12864-021-07527-z

BMC Genomics

RESEARCH ARTICLE Open Access

A highly mutable *GST* is essential for bract colouration in *Euphorbia pulcherrima* Willd. Ex Klotsch

Vinicius Vilperte[1,2], Robert Boehm[3] and Thomas Debener[1*]

Abstract

Background: Mutation breeding is an extraordinary tool in plant breeding to increase the genetic variability, where mutations in anthocyanin biosynthesis are targets to generate distinctive phenotypes in ornamental species. In poinsettia, ionizing radiation is routinely applied in breeding programs to obtaining a range of colours, with nearly all pink and white varieties being obtained after γ- or X-ray mutagenesis of red varieties. In the present study we performed a thorough characterization of a potential mutagenesis target gene as the main responsible for the '*white paradox*' in poinsettia.

Results: We identified a *GST* gene in poinsettia (*Bract1*) as an essential factor for the expression of anthocyanin-based red colouration of bracts, which presents a high phylogenetic similarity to known anthocyanin-related GSTs. Red poinsettia varieties and white mutants generated from these varieties by X-ray were analysed for polymorphisms related to the '*white paradox*' in the species. A 4 bp mutation in a short repeat within the coding region of *Bract1* is most likely responsible for the appearance of white phenotypes upon irradiation treatment. The polymorphism between wild-type and mutant alleles co-segregates with the phenotype in progeny from heterozygous red and white parents. Moreover, overexpression of *Bract1* wild-type allele in Arabidopsis *tt19* mutants restored the anthocyanin phenotype, while the *Bract1* mutated allele showed to be non-functional.

Conclusions: The identified repeat seems to be highly unstable, since mutated plants can be easily detected among fewer than 200 shoots derived from 10 mutated plants. Our data indicate that particular short repeat sequences, similar to microsatellite sequences or so-called dynamic mutations, might be hot spots for genetic variability. Moreover, the identification of the *Bract1* mutation fills a gap on the understanding on the molecular mechanism of colour formation in poinsettia.

Keywords: Anthocyanin, *Euphorbia pulcherrima*, Ionizing radiation, Glutathione S-transferase, Mutation breeding, Poinsettia, Short repeat sequences

Background

Poinsettia, *Euphorbia pulcherrima* Willd. ex Klotsch, commonly known as Christmas Star, is an important ornamental crop, especially due to its association with Christmas time in North America, Europe, and Asia, with annual sales reaching nearly 150 million dollars in the USA [69]. Its ornamental value is based on its intensely coloured bracts, which can be red, white, pink, or yellow or even have dual, scattered, or marbled colourations. Nonetheless, poinsettia breeding still focuses on red- and white-coloured varieties due to higher acceptance by consumers. In 2018, in Germany, approximately 80% of the poinsettias grown were red, 11% were

* Correspondence: debener@genetik.uni-hannover.de
[1]Institute of Plant Genetics, Leibniz Universität Hannover, 30419 Hannover, Germany
Full list of author information is available at the end of the article

white, and 9% were pink or had dual/scattered colouration [70].

Ionizing radiation is an important tool in mutation breeding for new colour variations in poinsettia, with nearly all pink and white varieties being obtained after gamma or X-ray mutagenesis of shoots of red varieties. Poinsettia mutation breeding is usually performed on cuttings that are irradiated with moderate dosages (~ 20 Gy), and mutants are selected on side shoots of the originally irradiated shoots. Flowering induction in the species occurs under short-day conditions and is accompanied by the development and colouration of bracts. Therefore, green leaves and red bracts occur concomitantly and accumulate different groups of pigments, i.e., chlorophylls and anthocyanins [53, 61]. Several anthocyanin types have been identified in poinsettia bracts and are responsible for its colouration range [3, 55, 66]; however, molecular information is still limited for the species [28, 72]. Nonetheless, genes responsible for the biosynthesis of the anthocyanin pathway have been intensively characterized in a range of species, with its regulation being highly dependent on R2R3-MYB regulatory genes and MYB-bHLH-WD40 (MBW) regulatory complexes [16, 58, 76].

Once synthesized on the cytoplasmic surface of the endoplasmic reticulum (ER), anthocyanin molecules need to be stored in the vacuole to prevent oxidation and loss of colour [4]. Two main models of anthocyanin transport have been proposed: i) a vesicle trafficking-mediated model, where vesicle-like structures filled with anthocyanins are imported into the central vacuole via vesicle fusion [23, 27, 62]; and ii) a transporter-mediated model, where anthocyanins are carried across the vacuolar membrane by transport proteins (e.g., ABC and MATE transporters) with the help of glutathione *S*-transferase (GST) enzymes [26, 63, 78]. GSTs can bind to anthocyanin molecules to form a complex, thus escorting them from the ER to the vacuole, preventing oxidation [13, 54, 67, 78]. Anthocyanin-related GSTs play major roles in anthocyanin transport, since loss of function of these proteins leads to phenotypes with a lack of pigmentation, such as *bz2* (*Bronze-2*) in maize, *an9* (*Anthocyanin 9*) in petunia, *tt19* (*Transparent Testa 19*) in Arabidopsis, *fl3* (*Flavonoid3*) in carnation, *riant* (*regulator involved in anthocyanin transport*) in peach, and *rap* (*reduced anthocyanin in petioles*) in strawberry [2, 8, 38, 43, 48]. In our previous study, an anthocyanin-related GST-like gene showed higher expression in a red poinsettia variety than in the white counterpart, thus making it a promising candidate responsible for the so-called *'white paradox'*, e.g. appearance of acyanic (uncolored) phenotype despite the detection of expression of all structural genes and the related enzyme activities involved in the formation of red anthocyanin pigments [72].

In our current study, we identified an anthocyanin-related *GST* as the most likely target of the radiation-induced mutation of red poinsettias in white bract sports. Using different approaches, this study demonstrates the functionality of the poinsettia *GST* as an anthocyanin transporter. Most importantly, we show that a short repeat motif within the coding region of the gene is highly unstable upon mutation treatment, which leads to the high frequency of anthocyanin mutations observed in commercial mutation breeding. In addition to facilitating mutation breeding for bract colours, these results may be a starting point for analysing the genetic instability of short repeat sequences in plants.

Results

Identification and characterization of *Bract1*

In a previous study [72], we observed higher expression of an anthocyanin-related *GST-like* gene (termed *Bract1* hereafter) in the red poinsettia variety 'Christmas Feelings' than in its white counterpart 'Christmas Feelings Pearl'. To investigate whether a similar phenomenon is observed in other red and white poinsettia pairs, we performed RT-qPCRs for six pairs of red-bracted poinsettia varieties and their independently generated white mutants. Normalized relative quantity (NRQ) values were calculated relative to one of the biological replicates of the 'Chr. Glory' variety according to the Pffafl method and equations [59]. The levels of *Bract1* expression varied among all varieties, with the varieties 'Christmas Feelings', 'Titan' and 'SK130' showing the highest relative expression. Although no lack of expression was observed in any of the white varieties, all red varieties showed significantly higher expression of *Bract1* than their white counterparts (Fig. 1).

To further characterize the anthocyanin-related *GST* in poinsettia we sequenced the complete coding and intronic regions of the gene for the 'Vintage' variety. The final full-length (from first ATG start codon to stop codon) *Bract1* sequence contains 2314 bp, with three exons (147 bp, 48 bp and 450 bp) and two introns (455 bp and 1214 bp) (Fig. 2a). The coding sequences (CDSs) of *Bract1* from 14 red- and white-bracted poinsettia varieties ('Noel', 'Valentino', 'Christmas Feelings', 'Christmas Feelings Pearl', 'Christmas Glory', 'Christmas Glory White', 'Joy', 'Joy White', 'Titan', 'Titan White', 'Bravo', 'Bravo White', 'SK130' and 'SK130 White') were further sequenced. The sequence alignment showed high similarity in the CDS for all varieties, except for six single-nucleotide polymorphisms (SNPs) that were identified in both the 'Chr. Glory' and 'Bravo' varieties (Table 1). This result shows the presence of at least two allelic forms of the *Bract1* gene. Additionally, a 4 bp deletion located 8 bp upstream of the first

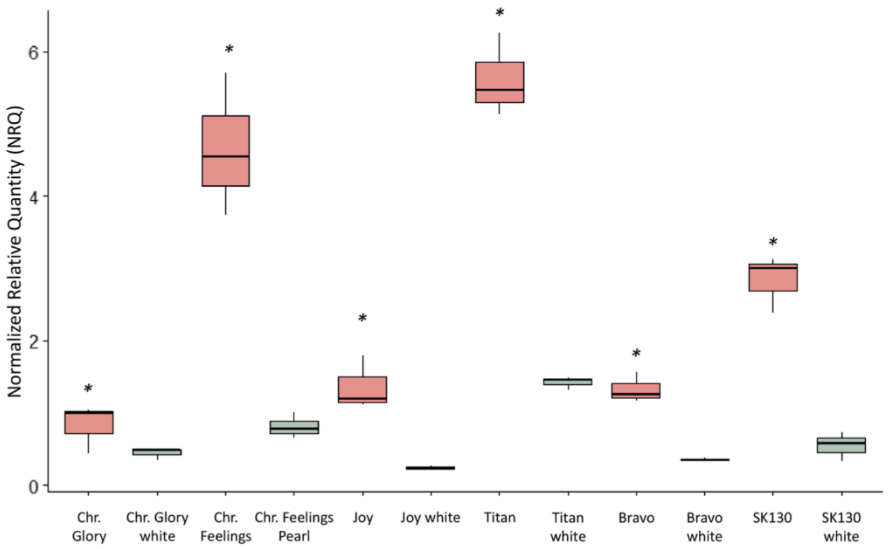


Fig. 1 RT-qPCR of *Bract1* for six pairs of red-bracted poinsettia varieties and their independently generated white mutants. The normalized relative quantity (NRQ) was calculated according to the Pfaffl equations [59] and using the 'Chr. Glory' variety as a reference sample. The '*' symbol indicates significant differences calculated with REST software between red and white pairs at $p \leq 0.05$

exon-intron junction was observed in all white varieties (Fig. 2b). The deletion is located in a short repeat locus, resembling a short simple sequence repeat (SSR), with a tetranucleotide motif ($(CTTC)_3$) composition. The exact location of the $(CTTC)_3$ motif is shown in Fig. 1a. The full-length gene sequence and CDS are available in Additional File S1.

Bract1 encodes a putative functional protein of 215 amino acids (aa) and a mass of 24.6 kDa, with distinctive GST components: a conserved GSH-binding site (G-site)

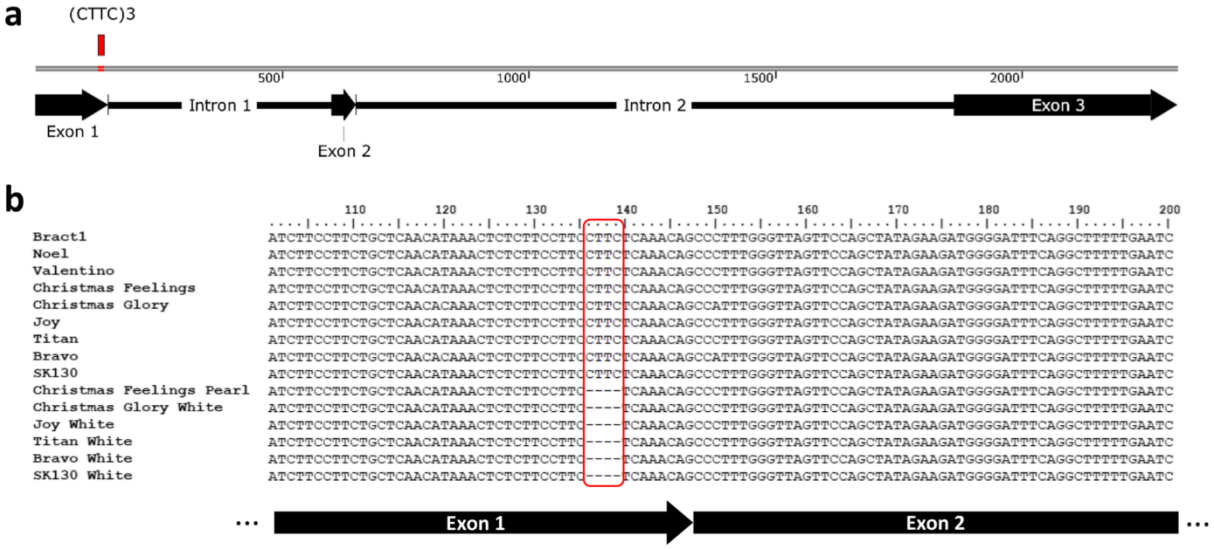


Fig. 2 Characterization of the anthocyanin-related GST gene (*Bract1*) in *Euphorbia pulcherrima*. **a** Schematic representation of the full-length sequence (2314 bp) of *Bract1* in the 'Vintage' variety. Black arrows represent the exonic regions. Black lines represent the intronic regions. The red square represents the location of the tetranucleotide motif SSR locus ($CTTC_3$). **b** Nucleotide alignment of the *Bract1* CDS for 14 red- and white-bracted poinsettia varieties. The figure shows a 100 bp region of the CDS in which a 4 bp deletion (red box) is observed only in the white varieties. Black arrows below the sequences show the location of the sequences in each exon. The first sequence corresponds to *Bract1* from the 'Vintage' variety and was used as a reference for the alignment

Table 1 List of SNPs identified in the sequenced varieties in comparison to the Vintage variety

Position	Original	Alternative	Varieties
90	T	A	Chr. Glory/Bravo
120	T	C	Chr. Glory/Bravo
150	C	A	Chr. Glory/Bravo
525	C	T	Chr. Glory/Bravo
578	A	G	Chr. Glory/Bravo
604	C	A	Chr. Glory/Bravo

located in the N-terminal domain and a C-terminal substrate-binding domain (H-site) [14]. The predicted protein from the CDS containing the 4 bp deletion is a putative truncated protein with an early stop codon at position aa52 due to a frameshift in the mRNA. The full-length amino acid sequence and the truncated version are available in Additional File S1.

Bract colouration associated with a deletion in the *Bract1* gene

The colour range in poinsettia varieties is obtained either through classic breeding (crossing) or mutagenic breeding (radiation), thus generating a spectrum of bract colours, such as pink, marble, orange and white/creamy. The white varieties are often obtained through radiation mutagenesis of the red varieties, followed by shoot development and trait selection. Therefore, red and white poinsettias from the same variety are referred to as 'pairs' due to their highly similar genetic background. However, not all red varieties can produce white sports through radiation. Therefore, red poinsettia varieties are distinguished into 'heterozygous' and 'homozygous' for the colouration locus according to their ability to generate white sports and according to the segregation of red and white phenotypes in progeny of crosses with white genotypes.

Since the 4 bp *indel* in the SSR locus of *Bract1* had shown indications of polymorphism among the different poinsettia varieties—and a correlation with bract colouration—we used a genotyping approach based on the fluorescent labelling of PCR fragments. We genotyped 22 different poinsettia varieties bearing red and white bracts (Fig. 3a, Additional File S2). All the red heterozygous varieties showed two distinct copies of the allele (with and without the 4 bp deletion), while their white counterparts showed only the copy with the deletion. On the other hand, homozygous red varieties (i.e., those unable to generate white sports) showed only the copy without the deletion.

We further genotyped a segregating population with 190 progeny from 'Joy' (Rr) x 'Joy White' (rr) containing 36 white and 154 red plants (Fig. 3b, Additional File S2). Contrary to expectation, we observed a deviation in the segregation ratio, which was approximately 4:1 (red:white), instead of the expected 1:1 ratio for this crossing. This may be explained by the fact that seeds from white varieties are less vital than those from red varieties (von Tubeuf, Selecta One, pers. comm.). In addition, white varieties also exhibit lower pollen fertility, thus increasing the chances of self-pollination when red varieties are used as a female parent (von Tubeuf, Selecta One, pers. comm.). In fact, 17 red progeny showed only the wild-type copy of the allele (data not shown), which can be attributed only to self-pollination. Nonetheless, all the white progeny showed only the allele copy containing the deletion, thus reinforcing our hypothesis that the presence of the allele containing the deletion in a homozygous recessive state is correlated with the white phenotype.

Bract1 is the anthocyanin-related GST orthologue in poinsettia

As GST genes occur in large gene families, we wanted to analyse whether the poinsettia GST gene was related to other GST genes involved in anthocyanin transport to

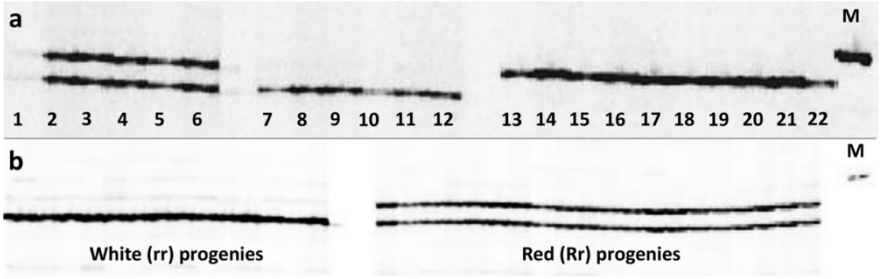


Fig. 3 PCR amplification of the tetranucleotide motif SSR locus $(CTTC)_3$ from the *Bract1* gene. **a** Band patterns from the amplified PCR fragments for *Bract1* in 22 red- and white-bracted poinsettia varieties. Samples 1–6 correspond to red heterozygous varieties, samples 7–12 correspond to white varieties, and samples 13–22 correspond to red homozygous varieties. **b** Example of the amplified PCR fragments for *Bract1* for the segregating population 'Joy' (Rr) x 'Joy White' (rr). M = marker. Figures were cropped for better visualization. Full length figures are available in Additional File S2

the vacuole. Therefore, we computed a phylogenetic tree from the deducted amino acid sequences of 95 GST family members from our previously assembled poinsettia transcriptome [72], as well as the *Bract1* and anthocyanin-related GSTs from other species (*CkmGST3*, *LcGST4*, *VvGST4*, *PhAN9*, *PpRiant1*, *PpRiant2*, *AtGSTF11* and *AtTT19*). Nine GST classes were identified among the poinsettia GSTs: Tau, Theta, Lambda, Zeta, Phi, tetrachlorohydroquinone dehalogenase (TCHQD), glutathionyl hydroquinone reductase (GHR), dehydroascorbate reductase (DHAR) and eukaryotic translation elongation factor 1B-γ (Ef1Bγ). Except Tau and Ef1Bγ, all other GST classes showed a single cluster (Fig. 4). All anthocyanin-related GSTs belong to the Phi class and clustered together in the phylogenetic tree, with *Bract1* showing high similarity with these GSTs.

By aligning the *Bract1* nucleotide CDSs with those of anthocyanin-related GSTs from other species, an overall nucleotide similarity of 61.9% was observed (Additional File S3). Protein alignment of BRACT1 with the other anthocyanin-related GSTs resulted in an overall similarity of 58.3%, with the peach RIANT1 protein showing the highest similarity (66.5%) (Additional File S3). Interestingly, we identified seven amino acid residues, previously reported as specific to anthocyanin-related GSTs [32, 37, 40], that are conserved in the protein alignment, except in AtGSTF11: 2Val, 11Ala, 13Cys, 62Phe, 90Leu, 91Glu and 152Ser (Fig. 5). In summary, these results indicate that *Bract1* is the anthocyanin-related GST orthologue in poinsettia.

Bract1 functionally complements the Arabidopsis *tt19* mutant phenotype

To examine the in vivo function of *Bract1* as an anthocyanin transporter, we tested the ability of *Bract1* cDNA to functionally complement the Arabidopsis GST mutant *tt19*, which is defective in the expression of

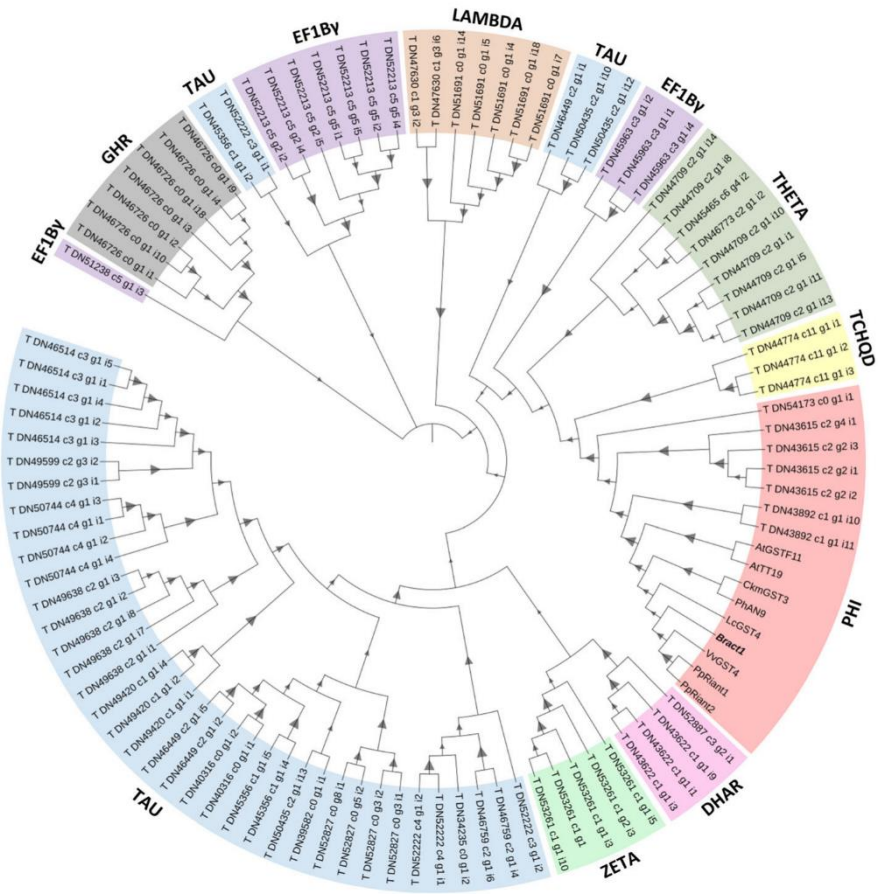


Fig. 4 Phylogenetic tree for 96 poinsettia GSTs and anthocyanin-related GSTs from other plant species. Amino acid sequences were aligned using MUSCLE. The maximum likelihood (ML) method based on the WAG matrix-based model was used to calculate the phylogenetic tree. Phylogenetic testing was performed using the bootstrap method with 1000 replicates, which are depicted as triangles, where the smallest value represents 1.3% and the largest 100%. Branch lengths were omitted for better visualization

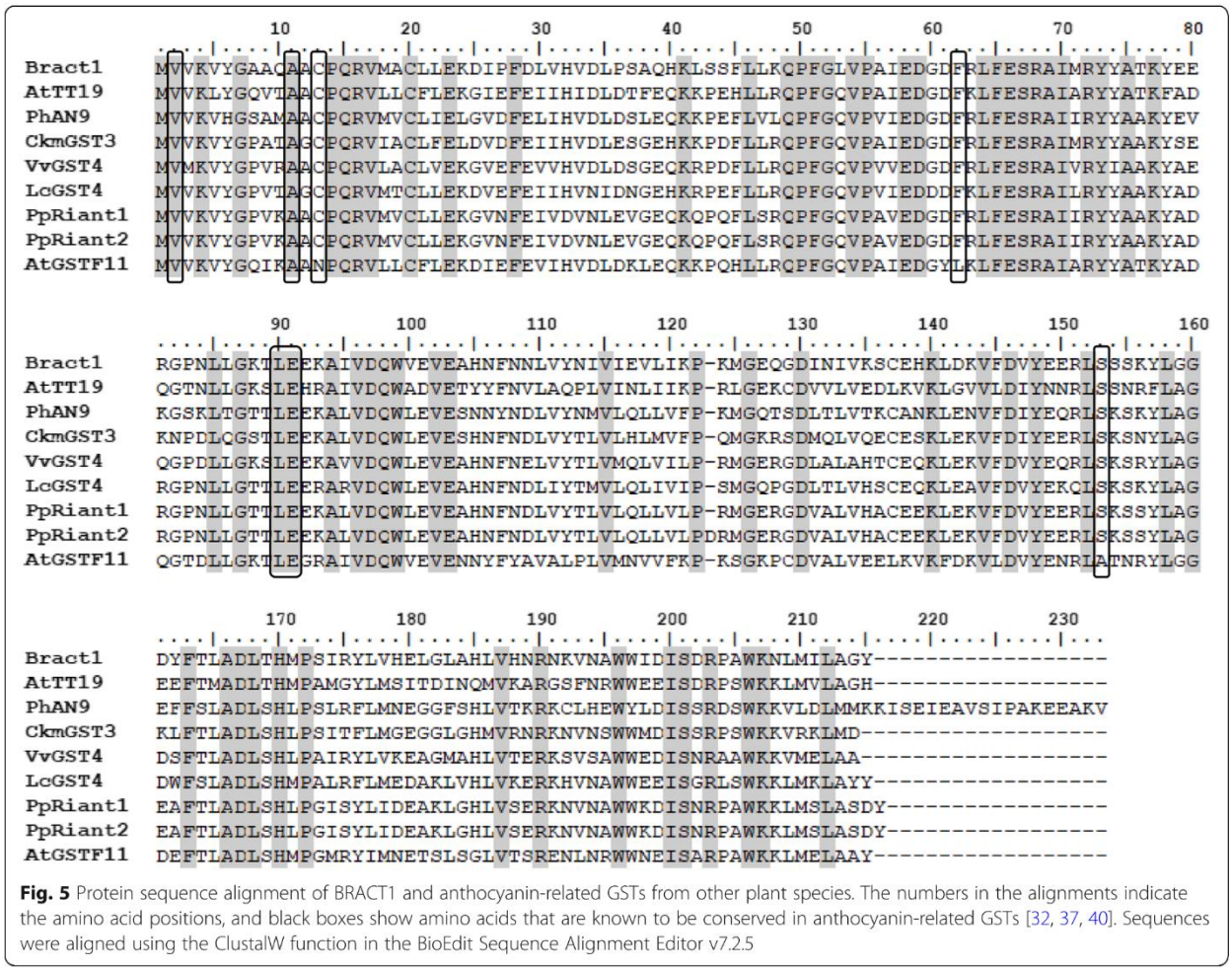


Fig. 5 Protein sequence alignment of BRACT1 and anthocyanin-related GSTs from other plant species. The numbers in the alignments indicate the amino acid positions, and black boxes show amino acids that are known to be conserved in anthocyanin-related GSTs [32, 37, 40]. Sequences were aligned using the ClustalW function in the BioEdit Sequence Alignment Editor v7.2.5

anthocyanins in aboveground organs and seeds. Two constructs containing the *Bract1* cDNA (with and without the 4 bp deletion) under the cauliflower mosaic virus (CaMV) 35S promoter were introduced into the *tt19* mutant by the floral-dip method [11, 75]. Although the constructs contained a GFP marker for the selection of transgenic events, we genotyped 10 independent biological replicates from the T_2 progeny of *tt19/35S::Bract1* and *tt19/35S::Bract1_mut* transgenic plants. All progeny contained the correct allele from the *Bract1* gene, thus confirming the correct integration of the transgenic construct (Fig. 6a, Additional File S2).

Upon stimulation of anthocyanin accumulation in seedlings by irradiation with red/blue LEDs, the *tt19/35S::Bract1* transgenic lines displayed a purple hypocotyl phenotype at the seedling stage, similar to the Columbia (*Col-0*) line but not the *tt19* mutant (Fig. 6b). On the other hand, *tt19/35S::Bract1* transgenic lines did not show complementation of the anthocyanin phenotype. The *ProAtUbi::GFP* construct, used as a control for infiltration, did not result in any phenotypic changes. Moreover, transgenic plants harbouring *Bract1* did not complement the seed colour of *tt19*, as the seed colour at the ripening stage remained the same as that of the mutant *tt19* in transgenic plants (Fig. 6c). This finding suggests that *Bract1* may have distinct functions from *TT19* during seed coat pigmentation. Taken together, these results not only emphasize the role of *Bract1* in anthocyanin transport in poinsettia but also demonstrate that a deletion in its coding region leads to a colourless phenotype.

De novo mutations occur with high frequency and include deletion of the 4 bp repeat

To study the stability of the 4 bp repeat within the first exon of the *Bract1* gene, we analysed DNA samples from mutation experiments conducted over the last 4 years at Selecta One. In brief, 10 cuttings from the varieties 'Aurora', 'SK159 Dark Pink', 'Aurora Jingle' and 'SK183' were X-ray irradiated with 20 Gy (30 Gy for 'SK183), and subsequently, side shoots from those cuttings were further propagated. DNA was extracted and analysed as

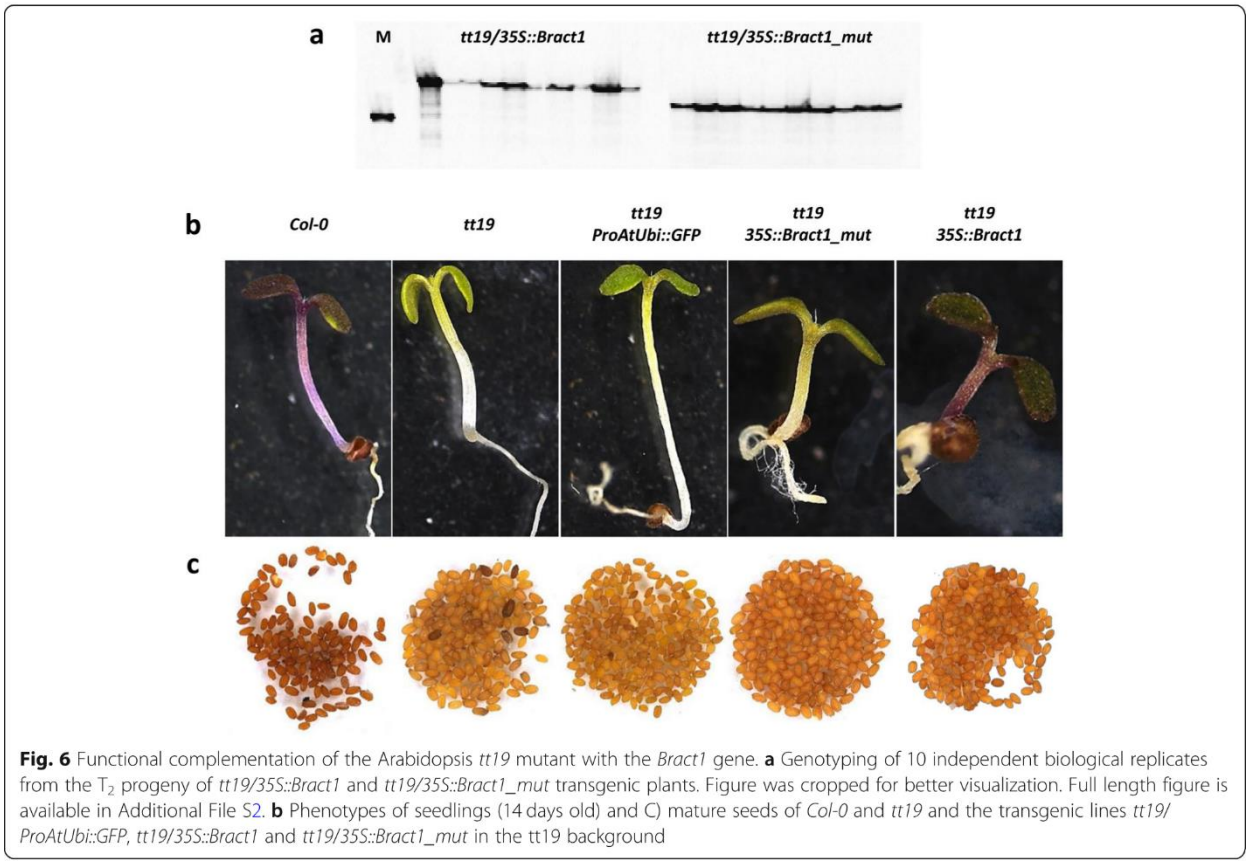


Fig. 6 Functional complementation of the Arabidopsis *tt19* mutant with the *Bract1* gene. **a** Genotyping of 10 independent biological replicates from the T_2 progeny of *tt19/35S::Bract1* and *tt19/35S::Bract1_mut* transgenic plants. Figure was cropped for better visualization. Full length figure is available in Additional File S2. **b** Phenotypes of seedlings (14 days old) and C) mature seeds of *Col-0* and *tt19* and the transgenic lines *tt19/ProAtUbi::GFP*, *tt19/35S::Bract1* and *tt19/35S::Bract1_mut* in the tt19 background

previously described from 377, 191, 188 and 186 of the propagated side shoots. Table 2 shows the results indicating that out of 942 samples, 9 mutations could be detected. Three mutated progeny were identified in both 'SK159 Dark Pink' and 'Aurora Jingle' individuals and two in the 'SK183' individuals, and only one mutated individual was identified in the 'Aurora' progeny. Unfortunately, as this was part of a commercial breeding programme, individual shoots were not labelled in a way that would allow tracing them back to one of the original shoots that were irradiated. However, even if all the mutations detected in each of the separate mutation treatments were redundant and originated from one original mutational event, the frequency was extraordinarily high.

Microsatellite repeats are not an anthocyanin-related feature

The microsatellite repeat present in the *Bract1* gene shows signs of instability upon irradiation treatment. To identify whether such repeats are a common feature for anthocyanin-related GSTs in Euphorbiaceae or related taxa or are a family-specific feature, we first computed a phylogenetic tree from the CDSs of *Bract1*, known anthocyanin-related GSTs (*CkmGST3*, *LcGST4*, *VvGST4*, *PhAN9*, *PpRiant1*, *PpRiant2*, and *AtTT19*) and GST-like orthologues from Euphorbiaceae species (*Euphorbia esula*, *Euphorbia pekinensis*, *Ricinus communis*, *Jatropha curcas*, *Hevea brasiliensis* and *Manihot esculenta*). Figure 7 shows that *Bract1* shared high similarity with the

Table 2 Fragment analysis of progeny from three X-ray-irradiated poinsettia varieties. Two methods were used for the fragment analysis: polyacrylamide gel electrophoresis (PAGE) and fragment length analysis (FLA) by capillary electrophoresis

Variety/year of irradiation	Number of progeny				Type of analysis
	Total	Homozygous (RR)	Heterozygous (Rr)	Homozygous (rr)	
Aurora/2016	377	376	1	0	PAGE
SK159 Dark Pink/2018	191	187	3	0	FLA
Aurora Jingle/2018	188	185	3	0	FLA
SK183/2018	186	1	184	1	FLA

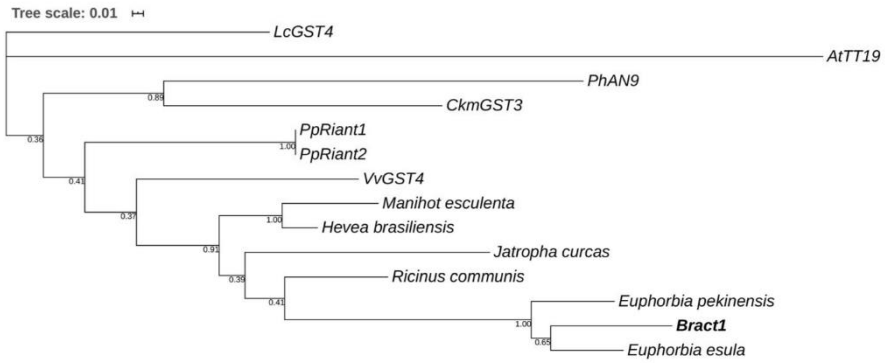


Fig. 7 Phylogenetic tree of anthocyanin-related GSTs and GST-like genes from Euphorbiaceae species. CDS nucleotide sequences were aligned using MUSCLE. The maximum likelihood (ML) method based on the T92 matrix-based model was used to calculate the phylogenetic tree. Phylogeny testing was performed using the bootstrap method with 1000 replicates, which are depicted next to the branches

GSTs from the two *Euphorbia* species (*E. esula* and *E. pekinensis)* but also closely clustered with the GSTs from the other Euphorbiaceae species. Although none of the GST-like genes from other Euphorbiaceae have been investigated as putative anthocyanin transporters, they may perform a similar function due to their homology with *Bract1* and the other anthocyanin-related GSTs.

Furthermore, we wanted to assess the distribution of the microsatellite repeat and possibly investigate its origin. The sequence alignment shows that none of the GSTs

```
Bract1                106 CCT TCT GCT CAA CAT AAA CTC TCT TCC TTC CTT CTC AAA CAG CCC TTT GGG TTA GTT CCA 165
                      36   Pro Ser Ala Gln His Lys Leu Ser Ser Phe Leu Leu Lys Gln Pro Phe Gly Leu Val Pro 55

Euphorbia esula       106 CAT TCT GCT CAA CAT AAA CTC CCT TCT TAC CTT TGC AAA CAG CCC TTT GGG TTA GTT CCA 165
                      36   His Ser Ala Gln His Lys Leu Pro Ser Tyr Leu Cys Lys Gln Pro Phe Gly Leu Val Pro 55

Euphorbia pekinensis  106 CAC TCT TCT CAC CAC AAA CTC CCT TCC TTT CTT CTC AAA CAG CCT TTT GGG CTC GTT CCG 165
                      36   His Ser Ser His His Lys Leu Pro Ser Phe Leu Leu Lys Gln Pro Phe Gly Leu Val Pro 55

Ricinus communis      106 GAT TCT GGA GAG CAT AAG CGA TCT GAA TTC CTT CTC AAA CAG CCA TTT GGG CAA GTC CCG 165
                      36   Asp Ser Gly Glu His Lys Arg Ser Glu Phe Leu Leu Lys Gln Pro Phe Gly Gln Val Pro 55

Jatropha curcas       106 AAA TCC GGG GAC CAT AAG CAT CCT GAC TTC CTT CTC AAA CAG CCG TTT GGG CAA GTT CCG 165
                      36   Lys Ser Gly Asp His Lys His Pro Asp Phe Leu Leu Lys Gln Pro Phe Gly Gln Val Pro 55

Hevea brasiliensis    106 GCT TCT GGA GAG CAC AAA CGA CCT GAC TTC CTT CTC AAA CAG CCA TTT GGG CAA GTT CCT 165
                      36   Ala Ser Gly Glu His Lys Arg Pro Asp Phe Leu Leu Lys Gln Pro Phe Gly Gln Val Pro 55

Manihot esculenta     106 GAT TCC GGC GAG CAA AAG CAA CCT CAC TTC CTT CTC AAA CAG CCA TTT GGC CAA GTT CCA 165
                      36   Asp Ser Gly Glu Gln Lys Gln Pro His Phe Leu Leu Lys Gln Pro Phe Gly Gln Val Pro 55

VvGST4                106 GAC TCT GGC GAG CAA AAA CGG CCT GAT TTC CTC CTT CGA CAG CCT TTT GGG CAA GTT CCA 165
                      36   Asp Ser Gly Glu Gln Lys Arg Pro Asp Phe Leu Leu Arg Gln Pro Phe Gly Gln Val Pro 55

PpRiant1              106 GAG GTG GGA GAG CAA AAG CAA CCT CAG TTC CTC TCC CGT CAG CCG TTT GGT CAA GTT CCA 165
                      36   Glu Val Gly Glu Gln Lys Gln Pro Gln Phe Leu Ser Arg Gln Pro Phe Gly Gln Val Pro 55

PpRiant2              106 GAG GTG GGA GAG CAA AAG CAA CCT CAG TTC CTC TCC CGT CAG CCG TTT GGT CAA GTT CCA 165
                      36   Glu Val Gly Glu Gln Lys Gln Pro Gln Phe Leu Ser Arg Gln Pro Phe Gly Gln Val Pro 55

CkmGST3               106 GAA TCC GGG GAG CAT AAG AAG CCC GAT TTT CTT CTT CGC CAG CCC TTT GGA CAA GTC CCA 165
                      36   Glu Ser Gly Glu His Lys Lys Pro Asp Phe Leu Leu Arg Gln Pro Phe Gly Gln Val Pro 55

PhAN9                 106 GAT TCT CTC GAG CAG AAA AAA CCT GAG TTT CTA GTT TTA CAG CCA TTT GGA CAA GTT CCT 165
                      36   Asp Ser Leu Glu Gln Lys Lys Pro Glu Phe Leu Val Leu Gln Pro Phe Gly Gln Val Pro 55

AtTT19                106 GAT ACA TTT GAG CAA AAA AAA CCA GAA CAT CTT CTT CGT CAG CCA TTT GGT CAA GTT CCA 165
                      36   Asp Thr Phe Glu Gln Lys Lys Pro Glu His Leu Leu Arg Gln Pro Phe Gly Gln Val Pro 55

LcGST4                106 GAC AAT GGA GAG CAT AAG CGT CCT GAG TTT CTT CTT CGA CAG CCT TTT GGG CAA GTT CCA 165
                      36   Asp Asn Gly Glu His Lys Arg Pro Glu Phe Leu Leu Arg Gln Pro Phe Gly Gln Val Pro 55
```

Fig. 8 Partial sequence alignment of *Bract1*, anthocyanin-related GSTs and orthologue GSTs from Euphorbiaceae species. The alignment spans a 60 bp region of the CDS containing the $(CTTC)_3$ SSR motif (represented in red)**.** The numbers in the alignments indicate the nucleotide and amino acid positions in the CDS. Sequences were aligned using the ClustalW function in the BioEdit Sequence Alignment Editor v7.2.5. The complete alignment is available in Additional File S7

analysed contain the same $(CTTC)_3$ SSR motif observed in *Bract1* (Fig. 8). When analysing the GSTs from Euphorbiaceae species, two CTTC repeat units were present in *E. pekinensis*, with a single nucleotide substitution responsible for the loss of the third repeat. *J. curcas*, *H. brasiliensis* and *M. esculenta* showed two CTTC repeat units, with two nucleotide substitutions related to the loss of the third repeat. Last, *E. esula* and *R. communis* showed a single CTTC unit and three nucleotide substitutions in the microsatellite region. The anthocyanin-related GSTs showed greater distinction in the microsatellite region. Three of the genes contained a single CTTC repeat unit (*CkmGST3*, *AtTT19* and *LcGST4*), while the others did not contain any repeat unit (*VvGST4*, *PpRiant1*, *PpRiant2* and *PhAN9*), which was due to the higher number of nucleotide substitutions (ranging from 4 to 7 substitutions). In conclusion, even though some of the known anthocyanin-related GSTs contain a CTTC sequence, CTTC repeats seem to be a common feature of Euphorbiaceae GSTs, and the three repeats from the *E. pulcherrima* GST (*Bract1*) behave in an unstable manner upon being subjected to ionizing irradiation.

Discussion

Anthocyanins, a class of flavonoid secondary metabolite compounds [47], are responsible for providing orange to blue colours in plant tissues, and their biosynthetic and regulatory mechanisms have been widely characterized [76]. However, there is still debate on the mechanism of anthocyanin transport from the ER to the vacuole [62, 63]. Strong evidence for the involvement of transport proteins [21, 24], with a special role of GST enzymes, has been reported in several plant species (Alfenito et al. [38, 43];). In the present study, we demonstrated that the *Bract1* gene functions as an anthocyanin transporter in poinsettia and that a highly mutable repeat in its coding region leads to frequent deletions and therefore to a colour-deficient phenotype.

Bract1 is a functional GST gene related to anthocyanin transport in poinsettia

GSTs are a large and diverse group of enzymes with multifunctional roles, especially in the detoxification of xenobiotics as well as in responses to biotic and abiotic stresses [1, 15]. The classification of GSTs is based on sequence conservation, genomic organization, and physiochemical properties, among other features [18, 33, 42]. Based on our previous study [72], we identified 95 *GST* genes in poinsettia and phylogenetically classified them into nine different classes based on their similarity with known Arabidopsis GSTs (Fig. 4). To date, 14 GST classes have been identified in plants: tau (U), phi (F), lambda (L), DHAR, theta (T), zeta (Z), EF1Bγ, TCHQD, microsomal prostaglandin E-synthase type 2 (mPGES-2), GHR, metaxin, Ure2p, hemerythrin (H) and iota (I) (reviewed by [42]).

A large number of GSTs have been identified in plant species, such as 49 in *Capsella rubella [31]*, 55 in Arabidopsis (), 61 in Citrus [45] and 139 in *L. chinensis [32]*. *Bract1* clusters with high bootstrap support with anthocyanin-related GSTs from other species (e.g., *AtTT19*, *PhAN9* and *VvGST4*), with all of these GSTs belonging to the phi class. Known anthocyanin-related GSTs belong almost exclusively to the phi class, except for *Bronze-2* from maize, which belongs to the tau class [50]. Further support for *Bract1* being a member of the phi-type plant GST genes is provided by the presence of two introns as a characteristic of this group of genes, such as *AN9* from petunia and *TT19* from Arabidopsis (Alfenito et al. [54];).

Complementation studies using Arabidopsis *tt19* mutants have been widely applied as proof of concept for the function of GSTs as anthocyanin transporters (Alfenito et al. [32, 34, 37, 40, 46, 54, 57];). Due to the high amino acid conservation of GST enzymes involved in flavonoid accumulation among species [77], they can complement each other's anthocyanin-deficient mutants (Alfenito et al. [43];). However, similar to our observation for *Bract1*, not all of these genes complemented both the shoot and seed phenotypes [34, 40, 48]. TT19 is involved in both anthocyanin accumulation in vegetative tissues and proanthocyanidin (PA) accumulation in Arabidopsis seed coats, which provides its brown colouration [38]. Transgenic *tt19* mutants overexpressing the petunia AN9 orthologue showed *AN9* mRNA expression in developing siliques, but the seed colour still remained the same as the wild-type mutant [38]. Altogether, these results suggest that GST orthologues from different species may have distinct functions from TT19 during seed-coat pigmentation.

A direct complementation of poinsettia white mutants with the functional *Bract1* would ultimately prove its function in bract colouration. However, neither *Agrobacterium*-mediated infiltration nor biolistic particle delivery system (a.k.a. gene gun) were successful for transient expression studies (data not shown). Stable transformation in poinsettia have been done using electrophoresis-based methods [9, 71], but no stable transgenic poinsettia was obtained. Successful stable transformation via *Agrobacterium*-mediated infiltration has been previously achieved, but the process is time-consuming [10]. Attempts to perform stable transformation of poinsettia with *Bract1* alleles will bridge the current knowledge gap but are out of the scope of the present study.

A loss-of-function mutation in *Bract1* is the cause of the "*white paradox*" in poinsettia

Based on our results, we hypothesize that deletion of one unit of the repeat in the *Bract1* gene is responsible

for most of the white genotypes in poinsettia. This hypothesis is strongly supported by the evidence that the *tt19/35S::Bract1_mut* mutant was not able to complement the anthocyanin phenotype in the Arabidopsis *tt19* mutant, unlike the *tt19/35S::Bract1* mutant. Mutations in GSTs leading to colourless phenotypes have been previously reported. A mutation in the *fl3* gene in carnation leads to a light pink phenotype, but a brighter phenotype is observed upon complementation by petunia *AN9* and maize *Bz2 [43]*. In peach, four alleles of a GST gene (*Riant*) were identified, with two of them containing frameshift mutations and unable to complement the Arabidopsis *tt19* phenotype. Varieties containing copies of the mutated alleles in a homozygous state showed flowers with white variegated phenotypes [8]. Last, a single-nucleotide polymorphism (SNP) in the strawberry *RAP* gene, leading to a premature stop codon, results in a mutant with green petioles and leaves. The non-functional *rap* gene was not able to complement Arabidopsis *tt19*, while wild-type *RAP* was successful [48].

In our analyses, all six independently generated white mutants of red varieties displayed the same deletion of a 4 bp repeat in *Bract1*, whereas the original varieties all contained a fully functional copy of the gene. In addition, co-segregation of the deletion with the white phenotype was observed in a segregated population of 190 progeny. Furthermore, a novel mutation leading to a homozygous recessive allele of *Bract1* among 184 samples obtained from irradiated cuttings of the heterozygous line SK183 led to a white phenotype (Table 3). Altogether, the results of this study present strong evidence that the four-base deletion in *Bract1* is the cause of the red-to-white shift in the poinsettia varieties analysed here. However, as anthocyanin biosynthesis involves several steps, other regulatory and structural genes might give rise to white mutants as well, as has been shown in numerous other examples [5, 39, 49, 52]. We did not detect these genes in our current plant material perhaps due to the much higher mutation rate of the *Bract1* gene than of less mutable genes.

Bract1 contains a short highly mutable four-base repeat

Upon X-ray treatment, red poinsettia plants produce progeny bearing white phenotypes with high frequencies, often based on only 10 irradiated cuttings (von Tubeuf, Selecta One, pers. comm., Selecta One). This phenomenon is associated with a deletion in a short repeat in the *Bract1* gene of white mutants in a homozygous state. The mutations in all six independent mutant pairs that we detected are exactly identical, which

Table 3 List of poinsettia varieties used in the present study

Variety ID	Variety name	Bract colour	Observation
1	Christmas Feelings	Red	
2	Christmas Glory	Red	
3	Joy	Red	
4	Bravo	Red	
5	Titan	Red	
6	SK130	Red	
7	Christmas Feelings Pearl	White	Mutation from Chr. Feelings
8	Christmas Glory White	White	Mutation from Chr. Glory
9	Joy White	White	Mutation from Joy
10	Bravo White	White	Mutation from Bravo
11	Titan White	White	Mutation from Titan
12	SK130 White	White	Mutation from SK130
13	Vintage	Red	
14	Christmas Aurora	Red	
15	Happy Day	Red	
16	Tabalunga	Red	
17	Christmas Day	Red	
18	Christmas Eve	Red	
19	Noel	Red	
20	Valentino	Red	
21	Prestige Red	Red	
22	Christmas Cracker	Red	

indicates that the X-ray treatment did not directly cause the mutation but rather led to changes indirectly by stimulating the DNA repair mechanisms via replication errors, by increasing recombination or by the other mechanisms discussed for mutations in repeat sequences [56]. The possible involvement of replication-based errors is supported by our observation that upon amplification of the repeat via standard PCR from cloned *Bract1* wild-type or mutant allele, a low level of variants carrying four-base indels can always be detected (data not shown).

Radiation is frequently used as a tool for mutagenic breeding in poinsettia. In contrast to ethylmethanesulphonate (EMS)-based chemical mutagenesis, which produces point mutations with high frequency [25], ionizing radiation (e.g., X-rays and γ-rays) induces DNA oxidative damage, such as double-strand breaks (DSBs), base substitutions, deletions and chromosomal alterations, at a lower frequency, frequently resulting in loss of gene function [35, 36, 51]. SSRs are among the most variable types of repetitive sequences in the genome [19]. Studies have shown that SSR instability increases with plant development [22] and abiotic stress [74]. This might be another explanation for the frequent observation of repeat changes in the *Bract1* gene after X-ray irradiation, although the small number of repeats (i.e., three) of four base pairs each does not fit the most widely applied criteria used to define SSRs, which usually focus on sequences with a larger number of repeats.

However, little information about the genetics and dynamics is available for short repeats. A majority of studies compared historical events for mostly shorter SSRs (2 and 3 bp repeats with larger repeat numbers) in present-day populations or the dynamic repeats responsible for human diseases (mostly trinucleotide repeats), which usually display effects beyond those of large numbers of repeats (> 30 repeats [56]).

Our observation that a large number of mutation events could be observed in the side shoots of ten irradiated plants indicates an unusually high mutation rate, which is in contrast to the few reports in which exact mutation rates have been reported for vegetatively propagated crops [65]. In one example, the woody ornamental plant *Tibouchina urvelliana* was irradiated three independent times with a 45 Gy dose, resulting in 0.06% dwarf mutants each time [65]. However, several authors reported that the radiosensitivity of vegetative tissues varies greatly among species and tissues [20], so exact comparative estimations of mutation frequencies have a very limited accuracy among species and conditions. However, experiments with transgenic Arabidopsis lines harbouring constructs designed to analyse restoration of GUS open reading frames by either recombination or by restoring in-frame translation by mutations in SSRs demonstrate the occurrence of easily detectable numbers of somatic mutation events [22, 74]. Together with the careful selection of side shoots after X-ray irradiation of poinsettia, this finding may explain the high rate of recessive mutations detected here.

In this study, we showed that the poinsettia *Bract1* gene is an active GST gene involved in the expression of anthocyanins in poinsettia bracts. Furthermore, a 4 bp deletion in a short repeat within the coding region of *Bract1* is the most likely cause of many mutations that lead to a white bract colour. This mutation occurs with an unusually high frequency and is presumably an indirect effect of X-ray mutagenesis. Future analyses using mutagenesis in transgenic *Arabidopsis* lines harbouring *Bract1* might help elucidate the causes of the high instability of this repeat. Moreover, this result might also serve as a reference for the study of other repeat-containing structural genes as potential mutational hot spots in plant genomes.

Conclusions

In this study, we showed that the poinsettia *Bract1* gene is an active GST gene involved in the expression of anthocyanins in poinsettia bracts. Furthermore, a four base pair deletion in a short repeat within the coding region of *Bract1* is the most likely cause of many white mutations for bract colour. This mutation occurs with an unusually high frequency and is presumably an indirect effect of the x-ray mutagenesis. Future analyses using mutagenesis in transgenic Arabidopsis lines harbouring the *Bract1* might help to elucidate the causes of the high instability of this repeat. Moreover, it might also serve as an example for other repeat containing structural genes in plant genomes as potential mutational hot spots.

Methods

Plant material

A range of red- and white-bracted varieties of poinsettia was used in this study for the different analyses (Table 3). Among the varieties, we used so-called pairs of red varieties and white mutants generated from these varieties by γ-ray or X-ray mutagenesis by the company Selecta One (Stuttgart, Germany). In addition, a segregating population containing 190 progeny from the "Joy x Joy white" cross, one such pair, was also used for analysis. The plant material used are exclusively cultivated varieties of the species *Euphorbia pulcherrima* which were provided by the company Selecta One. Identification of the material was conducted by Dr. Robert Boehm and Mr. Guido von Tubeuf. There were no vouchers taken and deposited.

Leaf and bract samples for DNA and RNA isolation were harvested at Selecta One, immediately frozen in liquid nitrogen, shipped on dry ice and stored at − 80 °C

at the Institute for Plant Genetics of the Leibniz Universität Hannover (Hannover, Germany). Moreover, rooted cuttings of three red ('Aurora', 'Aurora Jingle' and 'SK183') and one pink ('SK159 Dark Pink') variety were irradiated with γ-rays (20 Gy) and further developed, and new cuttings were generated by the company Selecta One. The progeny were used for fragment analysis (please refer to section "Poinsettia genotyping and fragment analysis").

Seeds of the *Arabidopsis thaliana* Columbia (*col-0)* genotype were available at the Institute for Plant Genetics of the Leibniz Universität Hannover; these seeds were originally obtained from the Arabidopsis information service in Frankfurt, Germany [12], and subsequently propagated in isolated greenhouse compartments. Seeds of the mutant line *tt19–8* (stock number: CS2105587) were obtained from the Arabidopsis Biological Resource Center (ABRC). Seeds were sown in Einheitserde P substrate, and seedlings were placed in long-day conditions (16 h light/8 h dark, 22 °C) for 2 weeks to induce flowering. Seedlings of the wild-type and mutants analysed for the expression of anthocyanins were grown under supplemental red/blue LED light (GP LED production DR/B 120 LB, Philips, Germany) to stimulate anthocyanin development.

DNA and RNA isolation

For the poinsettia samples, DNA was isolated from approximately 100 mg of leaf tissue using the NucleoSpin® Plant II Kit (Macherey–Nagel GmbH & Co., KG, Düren, Germany) according to the manufacturer's instructions. Total RNA was isolated from approximately 100 mg of bract tissue using the mirPremier™ miRNA Isolation Kit (Sigma-Aldrich, St. Louis, USA). For Arabidopsis samples, total RNA was isolated from approximately 50 mg of leaf tissue using the Quick-RNA Plant Kit (Zymo Research, Irvine, USA). cDNA synthesis was performed using the FastGene Scriptase Basic cDNA Kit (Nippon Genetics Europe GmbH, Düren, Germany) according to the manufacturer's recommendations. The DNA and total RNA concentrations and quality were assessed using a NanoDrop™ 2000 (Thermo Fisher Scientific, Wilmington, USA) and gel electrophoresis.

GST expression by RT-qPCR

Two endogenous reference genes (Translation elongation factor 1 beta (*EF1B*) and Translation elongation factor 1 alpha (*EF1A*)) were used to normalize the *Bract1* expression data. Primer sequences are available in Additional File S4. The amplification efficiency for all primers was obtained from relative standard curves. Three independent biological replicates were used for each of the varieties. 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 a 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 [59]. The real-time data for this study are provided according to the Minimum Information for Publication of Quantitative Real-Time PCR Experiments guidelines [6]. Pairwise statistical analysis between each red variety and its white counterpart was performed using the Relative Expression Software Tool (REST) v2.0.13 [60].

Bract1 gene sequencing

The poinsettia variety 'Vintage' was used for full-length sequencing of the GST-like gene (hereafter named *Bract1*). PCRs were performed in a 50 µL reaction containing 50 ng of DNA template, 1X PrimeSTAR® Buffer (Mg^{2+} plus), 0.2 mM each dNTP, 0.25 µM forward and reverse primers and 1.25 U of PrimeSTAR® HS DNA Polymerase (Takara Bio Inc., Kusatsu, Japan). The cycling conditions were 95 °C for 3 min; 30 cycles of 95 °C for 30 s, 60 °C for 30 s and 72 °C for 2 min; and a final extension of 10 min at 72 °C. The PCR products were resolved in a 1% (w/v) agarose gel by horizontal electrophoresis for 90 min at 100 V. The correct bands were excised from the gel and purified using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey–Nagel) following the manufacturer's recommendations. Finally, the purified PCR fragments were sent to Eurofins Genomics (Ebersberg, Germany) for Sanger sequencing. The generated sequences were aligned using the ClustalW function in the BioEdit Sequence Alignment Editor v7.2.5 [30], and a final full-length gene sequence for *Bract1* was generated.

The coding sequences (CDSs) of *Bract1* from 14 red- and white-bracted poinsettia varieties ('Noel', 'Valentino', 'Christmas Feelings', 'Christmas Feelings Pearl', 'Christmas Glory', 'Christmas Glory White', 'Joy', 'Joy White', 'Titan', 'Titan White', 'Bravo', 'Bravo White', 'SK130' and 'SK130 White') were further sequenced. PCRs were performed in a 50 µL reaction containing 1 µL of undiluted cDNA, 1X PrimeSTAR® Buffer (Mg^{2+} plus), 0.2 mM each dNTP, 0.25 µM forward and reverse primers and 1.25 U of PrimeSTAR® HS DNA Polymerase (Takara). The cycling conditions were 95 °C for 3 min; 30 cycles of 95 °C for 30 s, 60 °C for 30 s and 72 °C for 45 s; and a final extension of 10 min at 72 °C. The sequencing strategy was the same as that used for the full-length sequencing analysis. The generated sequences were aligned using the ClustalW

function in the BioEdit Sequence Alignment Editor v7.2.5. Primer sequences are available in Additional File S4.

Poinsettia genotyping and fragment analysis

To detect changes in the repeat structure of the *Bract1* gene, a genotyping approach based on the fluorescent labelling of PCR fragments [64] was applied. DNA samples were PCR amplified in a 20 µL reaction containing 50 ng of DNA template, 1X Williams buffer, 0.15 mM each dNTP, 0.0125 µM forward primer, 0.07 µM universal FAM-labelled M13 primer, 0.25 µM reverse primer and 1 U of DCSPol DNA Polymerase (DNA Cloning Service, Hamburg, Germany). The cycling conditions were 94 °C for 3 min; 24 cycles of 94 °C for 45 s, 59 °C for 1 min and 72 °C for 1 min; 6 cycles of 94 °C for 30 s, 52 °C for 45 s and 72 °C for 1 min; and a final extension of 10 min at 72 °C. Fifty microliters of formamide loading dye (98% formamide, 10 mM EDTA, 0.05% pararosanilin) was added to each reaction, which was then incubated at 95 °C for 5 min. The PCR products were resolved in a 6% (w/v) acrylamide gel via vertical electrophoresis using a LI-COR Gene Readir 4200 DNA Analyser (LI-COR Biosciences, Nebraska, USA). The varieties from Table 3 and the progeny of the irradiated variety 'Aurora' were genotyped using fluorescent labelling of PCR fragments. The progeny of the irradiated varieties 'Aurora Jingle' and 'SK159 Dark Pink' were analysed by capillary electrophoresis on an ABI 3730 XL system at Microsynth AG (Balgach, Switzerland). Primer sequences are available in Additional File S4.

Phylogenetic analysis

Protein sequences for *Bract1* and 95 different poinsettia GSTs, retrieved from our previous study [72], were predicted with TransDecoder [29] and used for the construction of a phylogenetic tree. Moreover, protein sequences for known anthocyanin-related GSTs from other species were included in the analysis: *CkmGST3* (*Cyclamen persicum* x *Cyclamen purpurascens*, GenBank - AB682678.1), *LcGST4* (*Litchi chinensis*, GenBank - KT946768.1), *VvGST4* (*Vitis vinifera*, GenBank - AY971515.1), *PhAN9* (*Petunia hybrida*, GenBank - Y07721.1), *PpRiant1* (*Prunus persica*, GenBank - KT312847.1), *PpRiant2* (*P. persica*, GenBank - KT312848.1), *AtGSTF11* (*Arabidopsis thaliana*, GenBank - NM_111189.3) and *AtTT19* (*A. thaliana*, GenBank - NM_121728.4). The putative protein sequences of all the GSTs are available in Additional File S5.

Sequence alignment was performed using MUSCLE [17], and the phylogenetic tree was constructed with MEGA X v10.0.5 [41] using the maximum likelihood (ML) method with the Whelan and Goldman matrix-based model using a discrete gamma distribution (WAG+G) [73]. The best model was estimated using MEGAX. The tree topology was tested via a bootstrap analysis with 1000 replicates. For better visualization of the phylogenetic tree, Tree Of Life (iTOL) software, version 4.2.3 [44] (https://itol.embl.de/), was used.

Plasmid construction and *Agrobacterium*-mediated infiltration

The coding sequence of *Bract1* was amplified from the poinsettia varieties 'Vintage' and 'Christmas Feelings Pearl' to capture both wild-type and mutated alleles. The primers used for amplification are available in Additional File S4. The PCR fragments were inserted in the sense orientation into the *BamHI-HindIII* site of the C757pGFPU10–35 s-ocs-LH (Additional File S6) binary vector (DNA Cloning Service, Hamburg, Germany) using the In-Fusion® HD Cloning Kit (Takara Bio Inc., Kusatsu, Japan). The vector contains a GFP gene under the control of the Arabidopsis ubiquitin promoter (*ProAtUbi::GFP*) and a 35S promoter upstream of the multiple cloning site into which inserts were cloned. The final expression vectors contained either the wild-type allele (*35S::Bract1*) or the mutated allele (*35S::Bract1_mut*). The vector containing only GFP was also used for transformation as a negative control (only *ProAtUbi::GFP*).

The expression vectors were introduced into *Agrobacterium tumefaciens* strain GV3101 via electroporation. Transformation of the *A. thaliana* mutant line *tt19–8* was performed using the floral dip method [11, 75]. For transgenic plant selection, T_0 seeds were sown in soil, and GFP-expressing seedlings were selected to produce T_1 and subsequently T_2 progeny to achieve *GST* homozygosity. T_2 seedlings of *tt19/35S::Bract1* and *tt19/35S::Bract1_mut* transgenic plants were used for phenotypic analysis. Non-transformed *tt19–8*, *Col-0* and *tt19/ProAtUbi::GFP* seedlings were used as controls. Seedlings were placed under red light to stimulate anthocyanin biosynthesis. To confirm correct *GST* integration, 10 independent biological replicates, each representing an independently selected transgenic line from the T_2 progeny of *tt19/35S::Bract1* and *tt19/35S::Bract1_mut* transgenic plants, were used for GST genotyping. The protocol was the same as that used in the section *"Poinsettia genotyping and fragment analysis"*.

Analysis of the GST repeat

To understand the origin of the microsatellite-like repeat in the *Bract1* gene, orthologous GST genes from Euphorbiaceae species were retrieved by BLASTN against the Euphorbiaceae (taxid: 3977) nucleotide database. GST-like genes from the Euphorbiaceae species *Ricinus communis* (GenBank - XM_002532928.3), *Manihot esculenta* (GenBank - XM_021748071.1), *Jatropha curcas* (GenBank - XM_012219312.2), *Hevea brasiliensis*

(GenBank - XM_021787187.1), *Euphorbia esula* (GenBank - PJAE01736713.1) and *Euphorbia pekinensis* [7, 72], as well the anthocyanin-related GSTs *CkmGST3* (*C. persicum* x *C. purpurascens*, GenBank - AB682678.1), *LcGST4* (*L. chinensis*, GenBank - KT946768.1), *VvGST4* (*V. vinifera*, GenBank - AY971515.1), *PhAN9* (*P. hybrida*, GenBank - Y07721.1), *PpRiant1* (*P. persica*, GenBank - KT312847.1), *PpRiant2* (*P. persica*, GenBank - KT312848.1) and *AtTT19* (*A. thaliana*, GenBank - NM_121728.4) were used to construct a phylogenetic tree.

Sequence alignment was performed using MUSCLE [17], and the phylogenetic tree was constructed with MEGA X v10.0.5 [41] using the ML method with the Tamura 3-parameter matrix-based model [68] using a discrete gamma distribution with invariant sites (T92 + G + I). The best model was estimated using MEGAX. The tree topology was tested via a bootstrap analysis with 1000 replicates. For better visualization of the phylogenetic tree, Tree Of Life (iTOL) software, version 4.2.3 [44] (https://itol.embl.de/), was used.

Abbreviations

ABC: ATP binding cassette transporter; CaMV: Cauliflower mosaic virus; cDNA: Complementary DNA; CDS: Coding sequence; DNA: Deoxyribonucleic acid; DSB: Double-strand break; EMS: ethyl-methanesulphonate; ER: Endoplasmic reticulum; GFP: Green fluorescent protein; GST: Glutathione S-transferase; GUS: β-glucuronidase protein; Gy: Gray (unit of ionizing radiation); Indel: Insertion and deletion variations; kDA: Kilo Dalton; MATE: Multi-antimicrobial extrusion protein; MBW complex: MYB-bHLH-WD40 complex; mRNA: Messenger RNA; NRQ: Normalized relative quantity; PCR: Polymerase chain reaction; RNA: Ribonucleic acid; RT-qPCR: Quantitative reverse transcription PCR; SNP: Single nucleotide polymorphism; SSR: Single sequence repeat; tt19: Transparent testa 19 Arabidopsis mutant

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07527-z.

Additional file 1 Full-length sequence of the *Bract1* gene, CDS sequences of red and white poinsettia varieties, full sequence and truncated version of the BRACT1 protein.

Additional file 2. Full length gel images referent to Fig. 3a, Fig. 3b and Fig. 6a from this publication.

Additional file 3 Sequence similarity of *Bract1* with other anthocyanin-related GSTs.

Additional file 4. List of primers used in each of the analyses in the present study.

Additional file 5 Deducted protein sequences from the *Bract1* and 95 GSTs from *E. pulcherrima* GSTs, as well as anthocyanin-related GSTs from other species.

Additional file 6 Schematic representation from the C757pGFPU10–35 s-ocs-LH binary vector used for the constructions of transformation plasmids containing either the wild-type allele (*35S::Bract1*) or the mutated allele (*35S::Bract1_mut*) from the poinsettia *GST*.

Additional file 7 Sequence alignment of *Bract1*, anthocyanin-related GSTs and orthologue GSTs from Euphorbiaceae species.

Acknowledgments

We would like to thank Guido von Tubeuf for providing the poinsettia varieties.

Authors' contributions

RB, TD and VV conceived the research. VV and TD designed the experiments. VV conducted the laboratory and bioinformatics analyses. RB provide the capillary electrophoresis results. VV and TD wrote the manuscript. All authors revised 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. Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. DNA sequence from the *Bract1* gene is available in the GenBank repository under accession number MW718861. The CDS sequences from several poinsettia genotypes generated during the current study are available in the GenBank repository under accession numbers MW718847 - MW718860.

Declarations

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

[1]Institute of Plant Genetics, Leibniz Universität Hannover, 30419 Hannover, Germany. [2]Present address: KWS SAAT SE & Co. KGaA, 37574 Einbeck, Germany. [3]Klemm + Sohn GmbH & Co., 70379 Stuttgart, KG, Germany.

Received: 29 September 2020 Accepted: 12 March 2021
Published online: 23 March 2021

References

1. Agrawal GK, Jwa N-S, Rakwal R. A pathogen-induced novel rice (Oryza sativa L.) gene encodes a putative protein homologous to type II glutathione S-transferases. Plant Sci. 2002;163(6):1153–60. https://doi.org/10.1016/S0168-9452(02)00331-X.
2. Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, et al. Functional Complementation of Anthocyanin Sequestration in the Vacuole by Widely Divergent Glutathione S-Transferases. Plant Cell. 1998;10(7):1135-49.
3. Asen S. Anthocyanins in bracts of Euphorbia Pulcherrima as revealed by paper chromatographic and spectrophotometric methods. Plant Physiol. 1958;33(1):14–7. https://doi.org/10.1104/pp.33.1.14.
4. Bassham DC. Pigments on the move. Nature. 2015;526(7575):644–5. https://doi.org/10.1038/526644a.
5. Ben-Simhon Z, Judeinstein S, Trainin T, Harel-Beja R, Bar-Ya'akov I, Borochov-Neori H, et al. A "white" anthocyanin-less pomegranate (Punica granatum L.) caused by an insertion in the coding region of the Leucoanthocyanidin Dioxygenase (LDOX; ANS) gene. PLOS ONE. 2015;10:e0142777. https://doi.org/10.1371/journal.pone.0142777.
6. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009;55(4):611–22. https://doi.org/10.1373/clinchem.2008.112797.
7. 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.

8. Cheng J, Liao L, Zhou H, Gu C, Wang L, Han Y. A small indel mutation in an anthocyanin transporter causes variegated colouration of peach flowers. J Exp Bot. 2015;66(22):7227–39. https://doi.org/10.1093/jxb/erv419.
9. Clarke JL, Klemsdal SS, Floistad E, Hvoslef-Eide AK, Haugslien S, Moe R, et al. Genetic engineering of poinsettia with the aim of enhancing its resistance to poinsettia mosaic virus. Acta Hortic. 2006:321–5. https://doi.org/10.17660/ActaHortic.2006.722.40.
10. Clarke JL, Spetz C, Haugslien S, Xing S, Dees MW, Moe R, Blystad DR. Agrobacterium tumefaciens-mediated transformation of poinsettia, Euphorbia pulcherrima, with virus-derived hairpin RNA constructs confers resistance to poinsettia mosaic virus. Plant Cell Rep. 2008;27(6):1027–38. https://doi.org/10.1007/s00299-008-0526-9.
11. Clough SJ, Bent AF. Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana: floral dip transformation of Arabidopsis. Plant J. 1998;16(6):735–43. https://doi.org/10.1046/j.1365-313x.1998.00343.x.
12. Debener T, Lehnackers H, Arnold M, Dangl JL. Identification and molecular mapping of a single Arabidopsis thaliana locus determining resistance to a phytopathogenic Pseudomonas syringae isolate. Plant J Cell Mol Biol. 1991;1(3):289–302. doi:10.1046/j.1365-313X.1991.t01-7-00999.x.
13. Dixon DP, Edwards R. Glutathione Transferases. Arab Book. 2010;8:e0131 doi: 10.1199/tab.0131.
14. Dixon DP, Lapthorn A, Edwards R. Plant glutathione transferases. Genome Biol. 2002;3, reviews3004.1. https://doi.org/10.1186/gb-2002-3-3-reviews3004.
15. Dixon DP, Skipsey M, Edwards R. Roles for glutathione transferases in plant secondary metabolism. Phytochemistry. 2010;71(4):338–50. https://doi.org/10.1016/j.phytochem.2009.12.012.
16. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15(10):573–81. https://doi.org/10.1016/j.tplants.2010.06.005.
17. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7. https://doi.org/10.1093/nar/gkh340.
18. Edwards R, Dixon DP, Walbot V. Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. Trends Plant Sci. 2000;5(5):193–8. https://doi.org/10.1016/s1360-1385(00)01601-0.
19. Ellegren H. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 2004;5(6):435–45. https://doi.org/10.1038/nrg1348.
20. Esnault M-A, Legue F, Chenal C. Ionizing radiation: advances in plant response. Environ Exp Bot. 2010;68(3):231–7. https://doi.org/10.1016/j.envexpbot.2010.01.007.
21. Francisco RM, Regalado A, Ageorges A, Burla BJ, Bassin B, Eisenach C, Zarrouk O, Vialet S, Marlin T, Chaves MM, Martinoia E, Nagy R. ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-Glucosides. Plant Cell. 2013;25(5):1840–54. https://doi.org/10.1105/tpc.112.102152.
22. Golubov A, Yao Y, Maheshwari P, Bilichak A, Boyko A, Belzile F, Kovalchuk I. Microsatellite instability in Arabidopsis increases with plant Development1[W][OA]. Plant Physiol. 2010;154(3):1415–27. https://doi.org/10.1104/pp.110.162933.
23. Gomez C, Conejero G, Torregrosa L, Cheynier V, Terrier N, Ageorges A. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST: anthocyanin trafficking in grapevine. Plant J. 2011;67(6):960–70. https://doi.org/10.1111/j.1365-313X.2011.04648.x.
24. Goodman CD, Casati P, Walbot V. A multidrug resistance–associated protein involved in anthocyanin transport in *Zea mays*. Plant Cell. 2004;16(7):1812–26. https://doi.org/10.1105/tpc.022574.
25. Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S. Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics. 2003;164(2):731–40.
26. Grotewold E, editor. The science of flavonoids. New York: Springer; 2006. https://doi.org/10.1007/978-0-387-28822-2.
27. Grotewold E, Davies K. Trafficking and sequestration of Anthocyanins. Nat Prod Commun. 2008;3, 1934578X0800300. https://doi.org/10.1177/1934578X0800300806.
28. Gu Z, Chen H, Yang R, Ran M. Identification of DFR as a promoter of anthocyanin accumulation in poinsettia (*Euphorbia pulcherrima* , willd. ex Klotzsch) bracts under short-day conditions. Sci Hortic. 2018;236:158–65. https://doi.org/10.1016/j.scienta.2018.03.032.
29. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512. https://doi.org/10.1038/nprot.2013.084.
30. Hall, T. A., Hall, T., Hall, T. A., Hall, T., Hall, T. A., Hall, T. A., et al. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Doi: 10.14601/Phytopathol_Mediterr-14998u1.29.
31. He G, Guan C-N, Chen Q-X, Gou X-J, Liu W, Zeng Q-Y, Lan T. Genome-wide analysis of the glutathione S-Transferase gene family in Capsella rubella: identification, expression, and biochemical functions. Front Plant Sci. 2016;7. https://doi.org/10.3389/fpls.2016.01325.
32. Hu B, Zhao J, Lai B, Qin Y, Wang H, Hu G. LcGST4 is an anthocyanin-related glutathione S-transferase gene in Litchi chinensis Sonn. Plant Cell Rep. 2016;35(4):831–43. https://doi.org/10.1007/s00299-015-1924-4.
33. Islam S, Sajib SD, Jui ZS, Arabia S, Islam T, Ghosh A. Genome-wide identification of glutathione S-transferase gene family in pepper, its classification, and expression profiling under different anatomical and environmental conditions. Sci Rep. 2019;9(1):9101. https://doi.org/10.1038/s41598-019-45320-x.
34. Jiang S, Chen M, He N, Chen X, Wang N, Sun Q, Zhang T, Xu H, Fang H, Wang Y, Zhang Z, Wu S, Chen X. MdGSTF6, activated by MdMYB1, plays an essential role in anthocyanin accumulation in apple. Hortic Res. 2019;6(1):40. https://doi.org/10.1038/s41438-019-0118-6.
35. Jo YD, Kim J-B. Frequency and Spectrum of radiation-induced mutations revealed by whole-genome sequencing analyses of plants. Quantum Beam Sci. 2019;3(2):7. https://doi.org/10.3390/qubs3020007.
36. Kazama Y, Hirano T, Saito H, Liu Y, Ohbu S, Hayashi Y, Abe T. Characterization of highly efficient heavy-ion mutagenesis in Arabidopsis thaliana. BMC Plant Biol. 2011;11(1):161. https://doi.org/10.1186/1471-2229-11-161.
37. Kitamura S, Akita Y, Ishizaka H, Narumi I, Tanaka A. Molecular characterization of an anthocyanin-related glutathione S-transferase gene in cyclamen. J Plant Physiol. 2012;169(6):636–42. https://doi.org/10.1016/j.jplph.2011.12.011.
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(1):104–14. https://doi.org/10.1046/j.1365-313X.2003.01943.x.
39. Koseki M, Goto K, Masuta C, Kanazawa A. The star-type color pattern in Petunia hybrida 'red star' flowers is induced by sequence-specific degradation of Chalcone synthase RNA. Plant Cell Physiol. 2005;46(11):1879–83. https://doi.org/10.1093/pcp/pci192.
40. Kou M, Liu Y, Li Z, Zhang Y, Tang W, Yan H, Wang X, Chen XG, Su ZX, Arisha MH, Li Q, Ma DF. A novel glutathione S-transferase gene from sweetpotato, IbGSTF4, is involved in anthocyanin sequestration. Plant Physiol Biochem. 2019;135:395–403. https://doi.org/10.1016/j.plaphy.2018.12.028.
41. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9. https://doi.org/10.1093/molbev/msy096.
42. Lallement P-A, Brouwer B, Keech O, Hecker A, Rouhier N. The still mysterious roles of cysteine-containing glutathione transferases in plants. Front Pharmacol. 2014;5. https://doi.org/10.3389/fphar.2014.00192.
43. 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(9):900–4. https://doi.org/10.1007/s00299-002-0545-x.
44. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(W1):W242–5. https://doi.org/10.1093/nar/gkw290.
45. Licciardello C, D'Agostino N, Traini A, Recupero G, Frusciante L, Chiusano M. Characterization of the glutathione S-transferase gene family through ESTs and expression analyses within common and pigmented cultivars of *Citrus sinensis* (L) Osbeck. BMC Plant Biol. 2014;14(1):39. https://doi.org/10.1186/1471-2229-14-39, Characterization of the glutathione S-transferase gene family through ESTs and expression analyses within common and pigmented cultivars of Citrus sinensis (L.) Osbeck.
46. Liu Y, Qi Y, Zhang A, Wu H, Liu Z, Ren X. Molecular cloning and functional characterization of AcGST1, an anthocyanin-related glutathione S-transferase

gene in kiwifruit (Actinidia chinensis). Plant Mol Biol. 2019;100(4-5):451–65. https://doi.org/10.1007/s11103-019-00870-6.
47. Liu Y, Tikunov Y, Schouten RE, Marcelis 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.
48. Luo H, Dai C, Li Y, Feng J, Liu Z, Kang C. Reduced Anthocyanins in petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. J Exp Bot. 2018;69(10):2595–608. https://doi.org/10.1093/jxb/ery096.
49. Luo P, Ning G, Wang Z, Shen Y, Jin H, Li P, et al. Disequilibrium of Flavonol synthase and Dihydroflavonol-4-Reductase expression associated tightly to white vs. red color flower formation in plants. *Front. Plant Sci.* 6. 2016. https://doi.org/10.3389/fpls.2015.01257.
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(6530):397–400. https://doi.org/10.1038/375397a0.
51. Morita R, Kusaba M, Iida S, Yamaguchi H, Nishio T, Nishimura M. Molecular characterization of mutations induced by gamma irradiation in rice. Genes Genet Syst. 2009;84(5):361–70. https://doi.org/10.1266/ggs.84.361.
52. Morita Y, Saito R, Ban Y, Tanikawa N, Kuchitsu K, Ando T, Yoshikawa M, Habu Y, Ozeki Y, Nakayama M. Tandemly arranged chalcone synthase a genes contribute to the spatially regulated expression of siRNA and the natural bicolor floral phenotype in Petunia hybrida: endogenous siRNA and bicolored flowers of Petunia. Plant J. 2012;70(5):739–49. https://doi.org/10.1111/j.1365-313X.2012.04908.x.
53. Moustaka J, Panteris E, Adamakis I-DS, Tanou G, Giannakoula A, Eleftheriou EP, Moustakas M. 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;154:44–55. https://doi.org/10.1016/j.envexpbot.2018.01.006.
54. Mueller LA, Goodman CD, Silady RA, Walbot V. AN9, a Petunia glutathione *S*-Transferase required for anthocyanin sequestration, Is a Flavonoid-Binding Protein. Plant Physiol. 2000;123(4):1561–70. https://doi.org/10.1104/pp.123.4.1561.
55. Nitarska D, Stefanini C, Haselmair-Gosch C, Miosic S, Walliser B, Mikulic-Petkovsek M, Regos I, Slatnar A, Debener T, Terefe-Ayana D, Vilperte V, Hadersdorfer J, Stich K, Halbwirth H. 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(1):216. https://doi.org/10.1186/s12870-018-1424-0.
56. Pearson CE, Edamura KN, Cleary JD. Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet. 2005;6(10):729–42. https://doi.org/10.1038/nrg1689.
57. Pérez-Díaz R, Madrid-Espinoza J, Salinas-Cornejo J, González-Villanueva E, Ruiz-Lara S. Differential roles for VviGST1, VviGST3, and VviGST4 in Proanthocyanidin and anthocyanin transport in Vitis vinifera. Front Plant Sci. 2016;7. https://doi.org/10.3389/fpls.2016.01166.
58. Petroni K, Tonelli C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Sci. 2011;181(3):219–29. https://doi.org/10.1016/j.plantsci.2011.05.009.
59. 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.
60. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 2002;30:10.
61. Pomar F, Ros Barceló A. Are red leaves photosynthetically active? Biol Plant. 2007;51(4):799–800. https://doi.org/10.1007/s10535-007-0164-z.
62. Poustka F, Irani NG, Feller A, Lu Y, Pourcel L, Frame K, Grotewold E. A trafficking pathway for Anthocyanins overlaps with the endoplasmic reticulum-to-vacuole protein-sorting route in Arabidopsis and contributes to the formation of vacuolar inclusions. Plant Physiol. 2007;145(4):1323–35. https://doi.org/10.1104/pp.107.105064.
63. Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR. The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. Plant Physiol Biochem. 2013;72:21–34. https://doi.org/10.1016/j.plaphy.2013.02.001.
64. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol. 2000;18(2):233–4. https://doi.org/10.1038/72708.
65. Schum, A., and Preil, W. (1998). "Induced Mutations in Ornamental Plants," in *Somaclonal Variation and Induced Mutations in Crop Improvement* Current Plant Science and Biotechnology in Agriculture. S. M. Jain, D. S. Brar, B. S. Ahloowalia (Dordrecht: Springer Netherlands), 333–366. doi:https://doi.org/10.1007/978-94-015-9125-6_17.
66. Slatnar A, Mikulic-Petkovsek M, Veberic 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.
67. Sun Y, Li H, Huang J-R. Arabidopsis TT19 functions as a carrier to transport anthocyanin from the cytosol to Tonoplasts. Mol Plant. 2012;5(2):387–400. https://doi.org/10.1093/mp/ssr110.
68. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Mol Biol Evol. 1992;9:678–87. https://doi.org/10.1093/oxfordjournals.molbev.a040752.
69. USDA (2018). Floriculture crops 2018 summary. National Agricultural Statistics Service. Available at: https://www.nass.usda.gov/Publications/Todays_Reports/reports/floran19.pdf.
70. van der Ploeg, R. (2018). FloraCulture international. Int. Assoc. Hortic. Prod. IAHP. Available at: https://www.floraculture.eu/digital-magazine/fci-november-december-2018/ [Accessed March 18, 2020].
71. Vik NI, Hvoslef-Eide AK, Gjerde H, Bakke K. Stable transformation of pionsettia via electrophoresis. Acta Hortic. 2001:101–3. https://doi.org/10.17660/ActaHortic.2001.560.12.
72. Vilperte V, Lucaciu CR, Halbwirth H, Boehm R, Rattei T, Debener T. Hybrid de novo transcriptome assembly of poinsettia (Euphorbia pulcherrima Willd. Ex Klotsch) bracts. BMC Genomics. 2019;20:900. https://doi.org/10.1186/s12864-019-6247-3.
73. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 2001;18(5):691–9. https://doi.org/10.1093/oxfordjournals.molbev.a003851.
74. Yao Y, Kovalchuk I. Abiotic stress leads to somatic and heritable changes in homologous recombination frequency, point mutation frequency and microsatellite stability in Arabidopsis plants. Mutat Res. 2011;707(1-2):61–6. https://doi.org/10.1016/j.mrfmmm.2010.12.013.
75. Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H. Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc. 2006;1(2):641–6. https://doi.org/10.1038/nprot.2006.97.
76. Zhao D, Tao J. Recent advances on the development and regulation of flower color in ornamental plants. Front Plant Sci. 2015;6. https://doi.org/10.3389/fpls.2015.00261.
77. Zhao J. Flavonoid transport mechanisms: how to go, and with whom. Trends Plant Sci. 2015;20(9):576–85. https://doi.org/10.1016/j.tplants.2015.06.007.
78. Zhao J, Dixon RA. The 'ins' and 'outs' of flavonoid transport. Trends Plant Sci. 2010;15(2):72–80. https://doi.org/10.1016/j.tplants.2009.11.006.

Publisher's Note

3.1 Supporting information

Additional File S1. Full-length sequence of the *Bract1* gene, CDS sequences of red and white poinsettia varieties, full sequence and truncated version of the *BRACT1* protein. (Available as electronic appendix).

Additional File S2. Full-length gel images referent to Figure 3a, Figure 3b and Figure 6a from this publication.

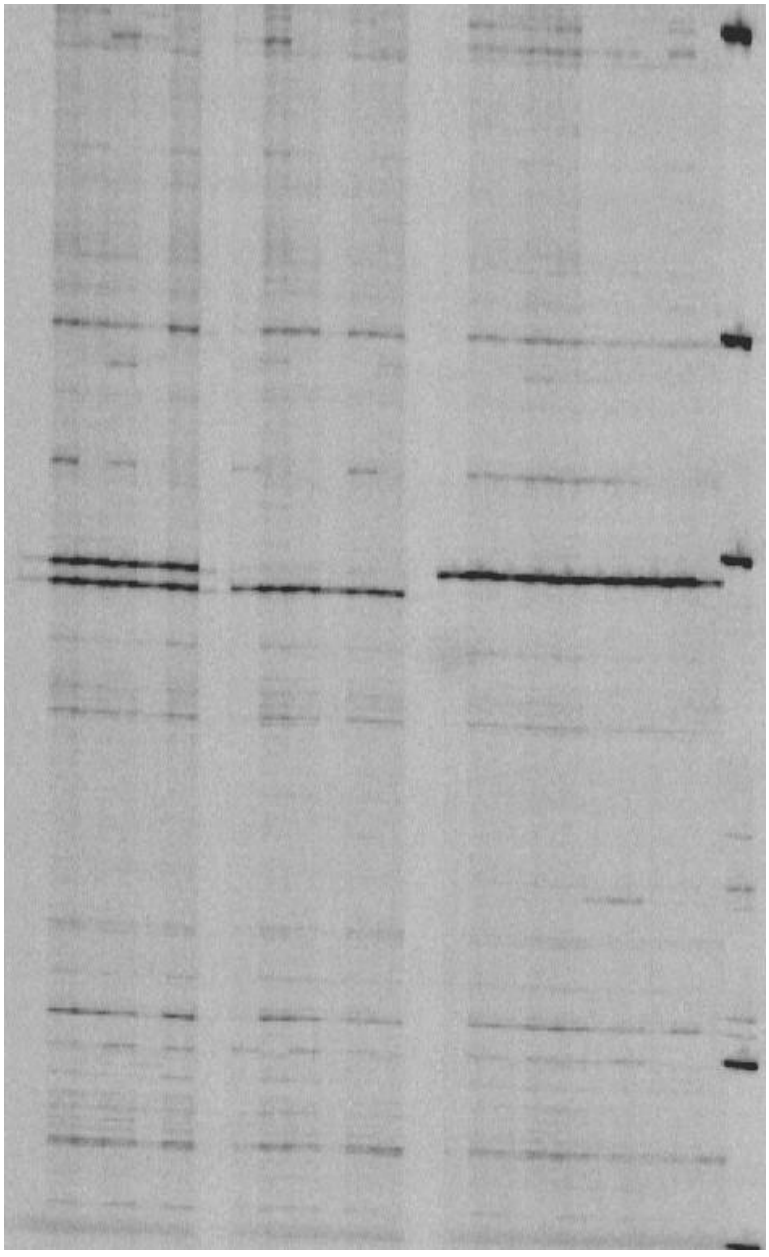


Figure 1. Full-length image referent to Figure 3a in this publication. Band patterns from the amplified PCR fragments for *Bract1* in 22 red- and white-bracted poinsettia varieties. Sample order is the same as presented in Figure 3a from the publication. The last lane on the right side corresponds to DNA marker for size estimation. The contrast in the image displayed in the manuscript was adjusted to remove some of the grey background. The alteration does not alter the interpretation of the results.

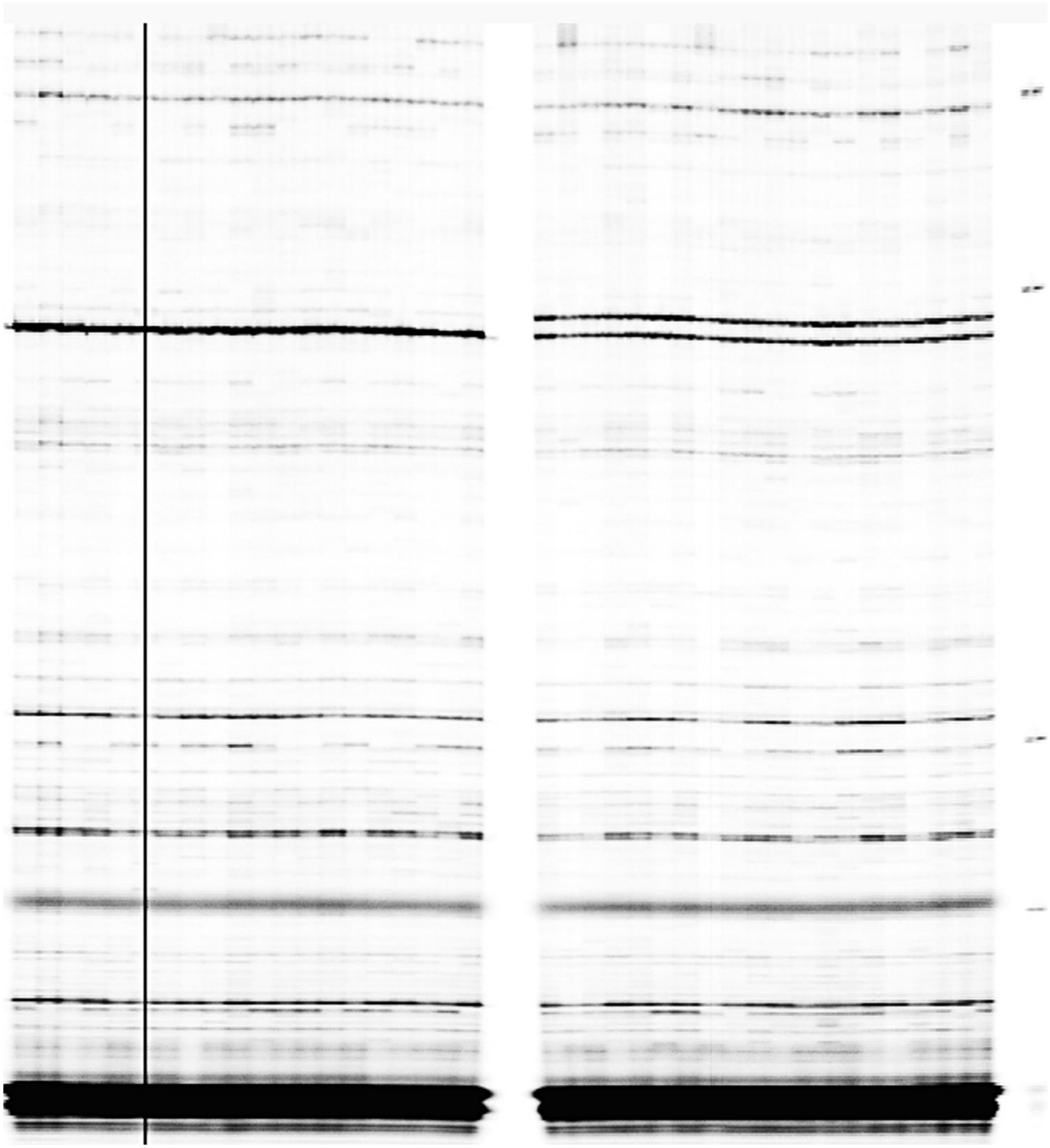


Figure 2. Full-length image referent to Figure 3b in this publication. Example of the amplified PCR fragments for *Bract1* for the segregating population 'Joy' (Rr) x 'Joy White' (rr). Sample order is the same as presented in Figure 3b from the publication. Last lane on the right side corresponds to DNA marker for size estimation. The black line seen on the left side of the image is due to a problem in the image capturing step. This was removed from the image displayed in the manuscript, but it does not alter the interpretation of the results. The samples cropped on the left side of the line also correspond to white (rr) progenies.

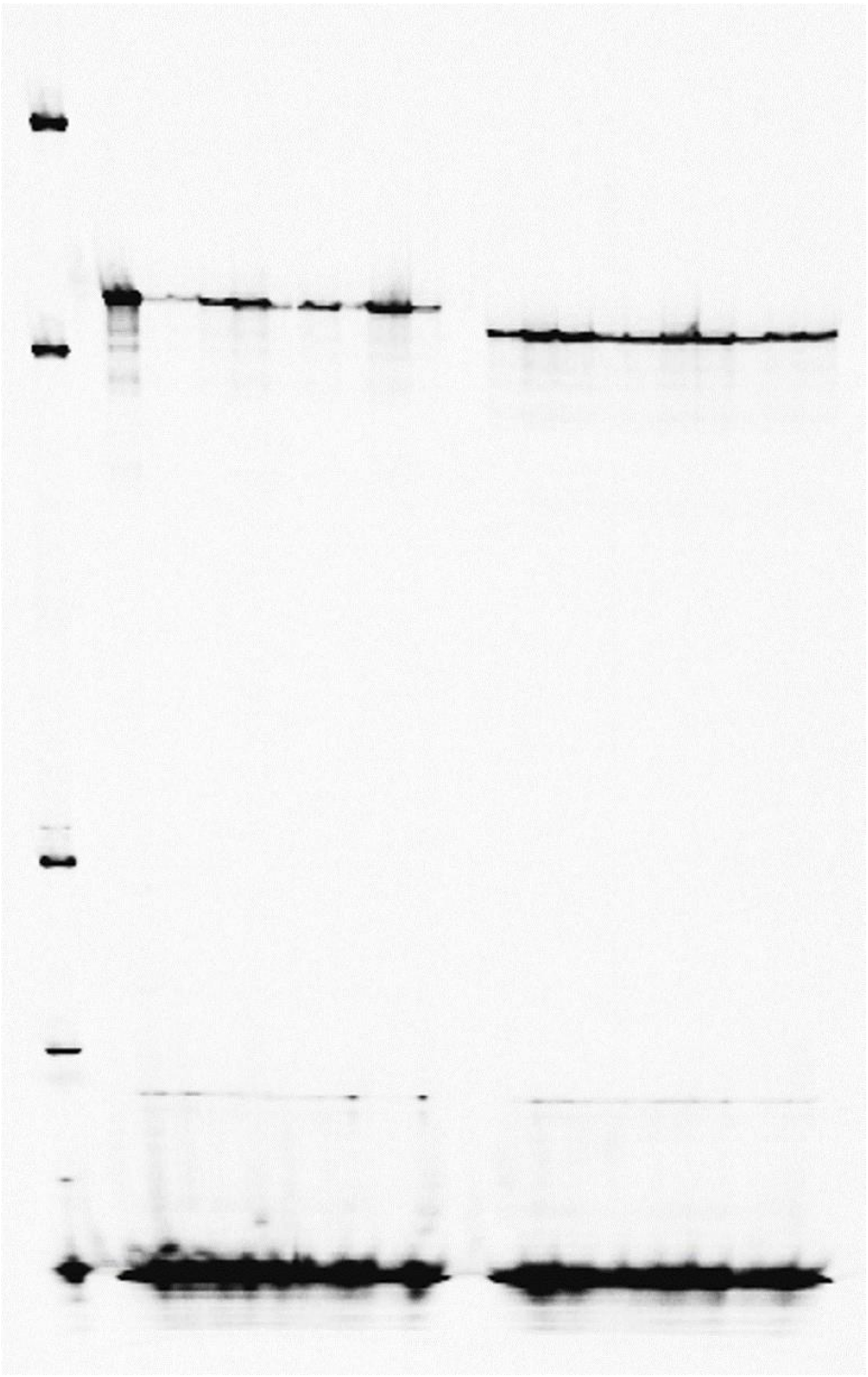


Figure 3. Full-length image referent to Figure 6a in this publication. Genotyping of 10 independent biological replicates from the T2 progeny of *tt19/35S::Bract1* and *tt19/35S::Bract1_mut* transgenic plants. Sample order is the same as presented in Figure 6a from the publication. The image in the publication was cropped to remove part of the image that does not contain any information and to be able to create a composite figure in the publication. The alteration does not alter the interpretation of the results.

Additional File S3. Sequence similarity of *Bract1* with other anthocyanin-related GSTs.

Genes	*Bract1*	AtTT19	PhAN9	CkmGST3	VvGST4	LcGST4	Riant1	Riant2	AtGSTF11
Bract1		60,4%	60,8%	63,2%	66,8%	65,5%	68,2%	49,6%	60,7%
AtTT19	55,1%		52,2%	58,9%	62,4%	62,7%	62,1%	50,5%	75,3%
PhAN9	53,8%	46,1%		63,6%	61,9%	60,3%	61,8%	46,7%	53,3%
CkmGST3	63,5%	52,8%	64,6%		67,1%	69,3%	69,4%	50,8%	59,6%
VvGST4	65,8%	57,0%	58,6%	68,0%		70,6%	74,8%	53,7%	62,3%
LcGST4	62,6%	57,9%	59,9%	66,8%	69,6%		71,7%	53,6%	61,0%
PpRiant1	66,5%	56,2%	59,9%	70,6%	75,8%	72,5%		67,8%	61,5%
PpRiant2	40,7%	37,9%	40,0%	44,4%	45,3%	46,2%	58,7%		47,3%
AtGSTF11	58,4%	68,6%	47,8%	53,7%	58,8%	56,5%	57,6%	37,0%	

Nucleotide sequence similarity

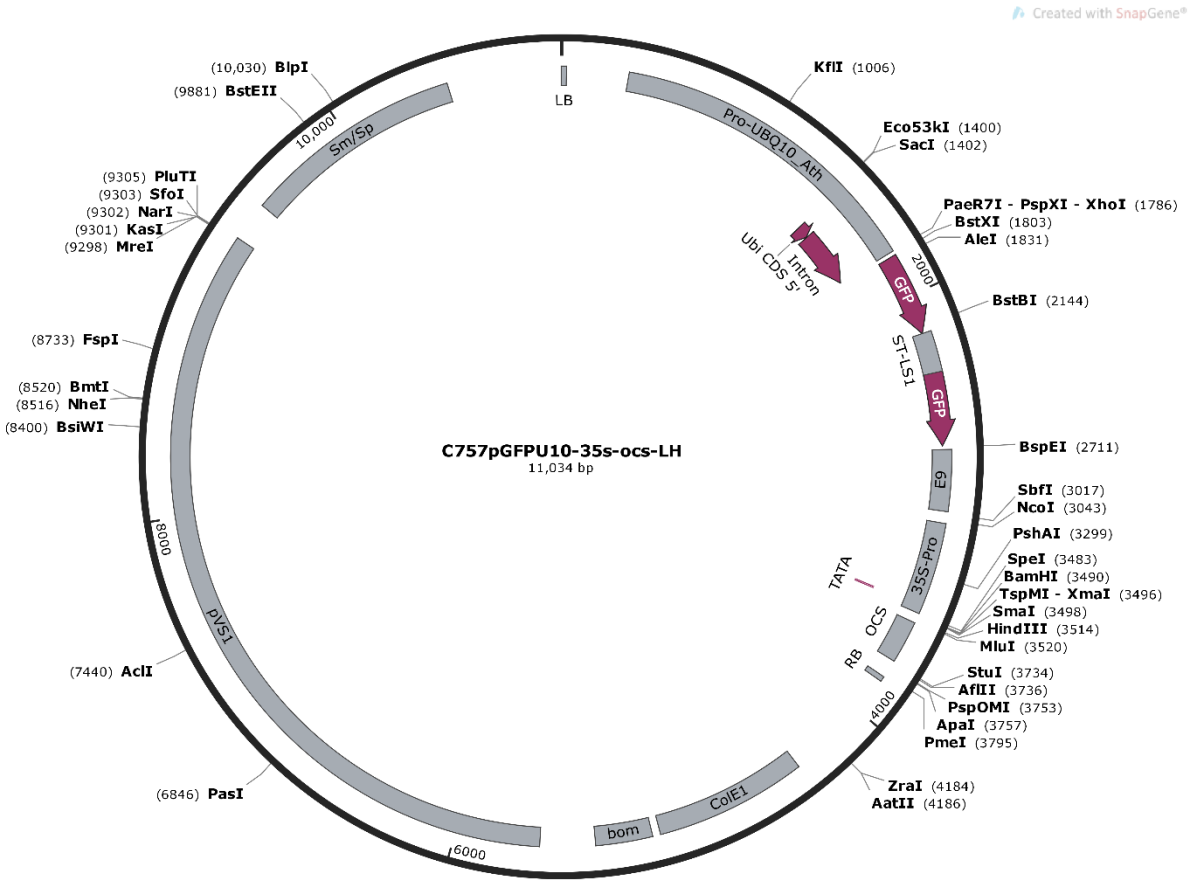
Protein sequence similarity

Additional File S4. List of primers used in each of the analyses in the present study.

Primer name	Sequence	Fragment size (bp)	Target sequence
	Sanger sequencing		
EpGST_F1	TCCGATCTAAGAAATCAAGGCTA	2,368 / 695	Full-length gene (gDNA) / full-length CDS (cDNA)
EpGST_R1	CAGTCGGCCGCTACATAGAT		
EpGST_F2	TGGCCTGCCTTTTAGAGAAA	596	1st intron, 1st exon (partially), 2nd exon (partially)
EpGST_R2	AAAGCCTGAAATCCCCATCT		
EpGST_F2	TGGCCTGCCTTTTAGAGAAA	1,933	2nd intron, 2nd exon (partially), 3rd exon (partially)
EpGST_R3	TATGGGCTTCCACTTCAACC		
	Genotyping		
EpGST_F3	GTAAAACGACGGCCAGTTGGCCTGCCTTTTAGAGAAA	197	$CTTC_3$ SSR locus in the 1st exon
EpGST_R4	ACAAGTTCAGGGGGCTGAG		
FAM labelled M13	FAM-GTAAAACGACGGCCAGT		
	RT-qPCR		
EpGST_qPCR_F	CTTGTTCATGAGCTTGGGTTAG	~ 80	*EpGST*
EpGST_qPCR_R	CGGTCGGTCCGATATATCAAT		
EF1B_qPCR_F	GGAGATGAGACTGAGGAGGATAA	~ 120	Translation elongation factor 1 beta – *EF1B*
EF1B_qPCR_R	CCCACGGCTTTACATCCAATA		
EF1A_qPCR_F	AAGATGATTCCCACCAAGCCCA	~ 120	Translation elongation factor 1 alpha – *EF1A*
EF1A_qPCR_R	CACAGCAAAACGACCCAGAGGA		
	InFusion Cloning		
Ep_infusion_GST_F	GACACTAGTTGGATCTCCGATCTAAGAAATCAAGGCTA	725	Full-length CDS (cDNA)
Ep_infusion_GST_R	CGACACGCGTAAGCTCAGTCGGCCGCTACATAGAT		

Additional File S5. Deducted protein sequences from the *Bract1* and 95 GSTs from *E. pulcherrima* GSTs, as well as anthocyanin-related GSTs from other species. (Available as electronic appendix).

Additional File S6. Schematic representation from the C757pGFPU10-35s-ocs-LH binary vector used for the constructions of transformation plasmids containing either the wild-type allele (*35S::Bract1*) or the mutated allele (*35S::Bract1_mut*) from the poinsettia *GST*.



Additional File S7. Sequence alignment of *Bract1*, anthocyanin-related GSTs and orthologue GSTs from Euphorbiaceae species. (Available as electronic appendix).

4 Development of a multiplex amplicon-sequencing assay to detect low-frequency mutations in poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) breeding programmes

Vinicius Vilperte [1], Robert Boehm [2] and Thomas Debener [1]

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

[2] Klemm + Sohn GmbH & Co., KG, 70379 Stuttgart, Germany

Type of authorship:	First author
Type of article:	Research article
Contribution to the article:	Planned and performed most of the experiments, analysed the data, prepared all tables and figures, and wrote most of the manuscript.
Contribution of the other authors:	Robert Boehm and Thomas Debener helped in conceptualizing the project and reviewing the manuscript. Thomas Debener contributed to the planning of the experimental design, data analysis, and writing and reviewing of the manuscript.
Supplements:	Supplementary material mentioned in the manuscript is provided as an electronic appendix.

Journal:	Plant Breeding
Date of publication:	24.05.2021
Impact factor:	1.662
DOI:	https://doi.org/10.1111/pbr.12925

Received: 23 November 2020 | Revised: 8 April 2021 | Accepted: 12 April 2021

DOI: 10.1111/pbr.12925

ORIGINAL ARTICLE

Plant Breeding WILEY

Development of a multiplex amplicon-sequencing assay to detect low-frequency mutations in poinsettia (*Euphorbia pulcherrima*) breeding programmes

Vinicius Vilperte[1] | Robert Boehm[2] | Thomas Debener[1]

[1]Institute of Plant Genetics, Leibniz Universität Hannover, Hannover, Germany

[2]Klemm+Sohn GmbH & Co., Stuttgart, Germany

Correspondence
Thomas Debener, Institute of Plant Genetics, Leibniz Universität Hannover, 30419 Hanover, Germany.
Email: debener@genetik.uni-hannover.de

Funding information
Marie Skłodowska-Curie Actions, Grant/Award Number: 675657

Communicated by: Henryk Flachowsky

Abstract

Poinsettia is an economically important ornamental potted plant in which certain bract colour variants are often obtained by mutation breeding. Previously, in poinsettia, we identified *Bract1*, a *GST* gene involved in the sequestration and transport of anthocyanins to the vacuole. This gene carries a short, highly mutable 4-bp repeat in its coding region. Loss of one repeat unit leads to a loss of function for *Bract1*, and in homozygous mutants, anthocyanin-based coloration is absent, resulting in white or cream-coloured bracts. Although mutation induction through ionizing radiation leads to a high frequency of mutations in *Bract1*, mutants are difficult to obtain from homozygous red genotypes. In this study, we used *Bract1*-specific amplicon sequencing as a tool to identify mutations in pools of tissues, which enabled the detection of mutations in dilutions of up to one mutant in 50 nonmutated samples. This approach enabled efficient screening of recalcitrant homozygous genotypes for mutated alleles and the reduction of the mutation load in the application of ionizing radiation in mutation breeding programmes.

KEYWORDS

amplicon-seq, *Bract1*, heterozygous mutations, irradiation mutagenesis, mutation breeding, poinsettia breeding

1 | INTRODUCTION

Mutation breeding is an established and successful method for generating variation in ornamental plants, because most economically important traits, for example, flower characteristics or growth habits, are easily monitored after mutagenic treatment (Ibrahim et al., 2018; Schum & Preil, 1998). Induced mutations in ornamental plants focus on economically valuable aesthetic values, such as flower and leaf characteristics (colour, size, morphology and fragrance), growth habits (compact, climbing and branching), physiological traits (photoperiodic response, early flowering, free flowering and flower keeping quality), and biotic and abiotic stress responses (Schum & Preil, 1998). The applicability of induced mutagenesis has been exploited in several ornamental species (Datta, 2020), with over 700 mutant varieties being obtained for more than 80 different species (FAO & IAEA, 2019). Among all ornamental-related traits, the development of new colours has strong appeal among consumers; therefore, it is one of the most exploited traits in mutagenesis breeding (reviewed by Datta, 2020; Ibrahim et al., 2018).

Poinsettia, also known as nochebuena, Weihnachtsstern or Christmas star (Trejo et al., 2012), is considered an important ornamental plant, especially due to its attractive, mostly red-coloured bracts, and it is primarily consumed during Christmas season in North America, Europe and Asia (Vilperte et al., 2019). The total consumption of ornamental poinsettias in Germany alone was estimated to be ~32 million potted plants in 2018, with ~90% of the sales comprising red- and white-coloured varieties (van der Ploeg, 2018). Nonetheless, the introduction of varieties with colour variations, such as pink, yellow and even bicoloured, scattered and marbled colourations, adds considerable ornamental value for product development. Ionizing radiation is an important tool in mutation breeding for obtaining such a range of colours, with nearly all pink and white varieties being obtained after γ- or X-ray mutagenesis of red varieties. Poinsettia mutation breeding is usually performed on cuttings that are irradiated with moderate dosages (~20–30 Gy), and mutants are selected on side shoots of the originally irradiated plants.

Anthocyanins have been identified as the main pigments in poinsettia bracts (Moustaka et al., 2018; Nitarska et al., 2018; Slatnar et al., 2013; Stewart et al., 1980), but extensive molecular studies of colour formation and accumulation remain lacking for this species. Dihydroflavonol-4-reductase (DFR) was suggested to promote the conversion of green leaves into red bracts (Gu et al., 2018). The orange colour of the poinsettia cultivar 'Harvest Orange' was related to a nonsense mutation in a flavonoid 3′-hydroxylase (*F3′H*) allele, thereby leading to a higher accumulation of pelargonidin (Nitarska et al., 2018). The pink colouration in pink and marble bracts can be attributed to periclinal chimeric structures (Nielsen et al., 2003). Previous unpublished studies, as well as the poinsettia transcriptome (Vilperte et al., 2019), have demonstrated the expression of most structural genes and the related enzyme activities involved in the formation of red anthocyanin pigments. Thus, the appearance of acyanic (uncoloured) varieties in poinsettias, despite gene expression being present, is referred to as the '*white paradox*'.

A glutathione *S*-transferase (GST) gene (*Bract1*) was recently determined to be involved in the transport and accumulation of anthocyanins in poinsettia bracts (Vilperte et al., 2019, 2020). A 4-bp deletion in a short repeat within the coding region is responsible for the lack of function of *Bract1* and thus explains the '*white paradox*' in poinsettia, that is, the appearance of white phenotypes upon irradiation treatment. The identified repeat seems to be highly unstable, because mutated plants can be easily detected among fewer than 200 shoots derived from an original set of only 10 mutated heterozygous plants (Vilperte et al.,2020). In contrast, de novo mutations derived from homozygous plants cannot be identified phenotypically due to the complete dominance of *Bract1* but only by molecular analysis.

In this study, we present a strategy for a multiplex amplicon-sequencing (hereafter named amplicon-seq) assay to detect newly generated heterozygotes among pooled progeny from irradiated poinsettia homozygotes and thus significantly increase the efficiency of the detection of heterozygous mutants. Employing different approaches, this study demonstrates the applicability of using an amplicon-seq strategy to detect low-frequency mutations in a heterozygous state, thereby enabling the detection of mutant alleles in larger pools of tissue samples. Most importantly, we developed a marker assay linked to colour variation that could be directly implemented in poinsettia breeding programmes, thereby contributing to the genetic resources by obtaining mutants from unexploited homozygous genotypes and improving the selection of prominent breeding material.

2 | MATERIALS AND METHODS

2.1 | Plant material

Three poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) varieties were utilized in this study: the red varieties 'Vintage' and 'Christmas Feelings' and the white variety 'Christmas Feelings Pearl'. The white variety 'Christmas Feelings Pearl' was obtained through γ-ray mutagenesis of the red variety 'Christmas Feelings' followed by shoot development and trait selection at the company Selecta One (Stuttgart, Germany; https://www.selecta-one.com/). Therefore, red- and white-bracted poinsettias from the same variety are referred to as 'pairs' due to their highly similar genetic backgrounds.

Rooted cuttings from all varieties were obtained from the company Selecta One and grown in a greenhouse at the Institute for Plant Genetics of the Leibniz Universität Hannover under short-day conditions to induce flower formation and to stimulate the development of coloured bracts. The plants were grown in 5-L pots containing Einheitserde P substrate (Hermann Meyer KG) with an average temperature of 22°C and 9 hr of daylight (15 hr of darkness).

2.2 | Plasmid controls

Plasmids obtained from Vilperte et al. (2020) were used as controls for amplicon sequencing. Briefly, the coding sequence of *Bract1* (GenBank—MW718861) was amplified from the poinsettia varieties 'Vintage' and 'Christmas Feelings Pearl' (GenBank—MW718847–MW718860) to capture both wild-type and mutated alleles (645/641 bp plus 50 bp of partial 5′ and 3′ flanking sequences) and inserted into the C757pGFPU10-35s-ocs-LH binary vector (DNA Cloning Service, Hamburg, Germany). The fragment was amplified using the primer pair EpGST_F1 (TCCGATCTAAGAAATCAAGGCTA) and EpGST_R1 (CAGTCGGCCGCTACATAGA; Vilperte et al., 2020). The vector contains a GFP gene under the control of the *Arabidopsis* ubiquitin promoter (*ProAtUbi::GFP*) and a 35S promoter upstream of the multiple cloning site into which inserts were cloned. The final expression vectors contained either the wild-type allele (*35S::Bract1*—695 bp) or the mutated allele (*35S::Bract1_mut*—691 bp).

2.3 | DNA isolation

DNA was isolated from ~100 mg of poinsettia leaf tissue using the NucleoSpin® Plant II Kit (Macherey–Nagel GmbH & Co.) according to the manufacturer's instructions. DNA concentrations and quality were assessed by photometry using a NanoDrop™ 2000 (Thermo Fisher Scientific) and by gel electrophoresis.

2.4 | Sample preparation

Plasmids containing the wild-type and mutated copies of the *Bract1* alleles were used as controls for the amplicon-seq analysis. Moreover, the poinsettia varieties were also used as controls: (a) 'Vintage' is homozygous dominant for the *Bract1* locus (designated RR), containing two copies of the wild-type allele; (b) 'Christmas Feelings' is heterozygous for the same locus (Rr) and contains both wild-type and mutated versions of the allele; and (c) 'Christmas Feelings Pearl' is homozygous recessive (rr) for the *Bract1* locus and contains two copies of the mutated allele. DNA samples from the varieties 'Vintage' and 'Christmas Feelings Pearl' were diluted to the same concentration (10 ng/μl) and used to create serial dilutions of wild-type (R) and mutated alleles (r). The following dilutions (R:r) were used: 1:1, 5:1, 10:1, 20:1, 50:1, 100:1 and 500:1.

2.5 | Genotyping

To detect changes in the repeat structure of the *Bract1* gene, a genotyping approach based on the fluorescent labelling of polymerase chain reaction (PCR) fragments (Schuelke, 2000) was employed. DNA samples were PCR-amplified in a 20-μl reaction containing 50 ng of DNA template, 1× Williams buffer, 0.15-mM each dNTP, 0.0125-μM forward primer, 0.07-μM universal IRD_{700}-labelled M13 primer, 0.25-μM reverse primer and 1 U of DCSPol DNA Polymerase (DNA Cloning Service, Hamburg, Germany). The cycling conditions were 94°C for 3 min; 24 cycles of 94°C for 45 s, 59°C for 1 min and 72°C for 1 min; six cycles of 94°C for 30 s, 52°C for 45 s and 72°C for 1 min; and a final extension of 10 min at 72°C. Fifty microlitres of formamide loading dye (98% formamide, 10-mM EDTA and 0.05% pararosaniline) was added to each reaction, which was subsequently incubated at 95°C for 5 min. The PCR products were resolved in a 6% (w/v) acrylamide gel via vertical electrophoresis using a LI-COR 4,200 DNA Analyser (LI-COR Biosciences). Primer sequences are available in Table 1.

2.6 | Amplicon sequencing

Control plasmids, poinsettia varieties 'Vintage', 'Christmas Feelingsa and 'Christmas Feelings Pearl', and serial dilutions of genomic DNA samples (1:1, 5:1, 10:1, 20:1, 50:1, 100:1 and 500:1) were submitted to amplicon sequencing. Briefly, *Bract1*-specific primers were tagged with an M13 tail and used in the first round of PCR (Table 2). Complementary M13 tags were subsequently used in a second PCR to differentiate PCR fragments arising from each variety. A schematic overview of the amplicon-seq approach is shown in Figure 1.

Both rounds of PCRs were performed in a 50-μl reaction containing 30 ng of DNA, 1× PrimeSTAR® Buffer (Mg^{2+} plus), 0.2 mM of each dNTP, 0.25 μM of forward and reverse primers and 1.25 U of PrimeSTAR® HS DNA Polymerase (Takara). The cycling conditions were 95°C for 3 min; 10 (for the second PCR) and 20 (for first PCR) cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 45 s; and a final extension of 10 min at 72°C. PCR products were pooled and sent to Eurofins Genomics Germany GmbH (Ebersberg) for Illumina sequencing. Primer sequences are available in Table 1.

Sequences were analysed using CLC Genomics Workbench (Qiagen). The Kruskal–Wallis test for non-parametric data ($p \leq .05$) was applied for pairwise comparisons of allele frequency. Statistical analysis and data plotting were performed by in-house scripts using the language and statistical environment R.

3 | RESULTS

3.1 | Detection of *Bract1* mutation by PCR fluorescent labelling

In a previous study (Vilperte et al., 2020), we used a genotyping approach based on the fluorescent labelling of PCR fragments to identify wild-type and mutated alleles in a range of poinsettia varieties. This approach produced satisfactory results to differentiate varieties containing a 4-bp *indel* in the *Bract1* locus in homozygous (dominant and recessive) and heterozygous states. However, heterozygous mutations arising from irradiation mutagenesis of homozygous dominant varieties have been shown to occur at low frequencies, and screening individual mutants by phenotype is impossible due to the dominance of *Bract1* (Vilperte et al., 2020). Therefore, we created serial dilutions of wild-type (R) and mutated (r) *Bract1* alleles to simulate mutation frequencies and evaluate the potential for detecting mutations at such dilutions through PCR-based techniques.

The amplification of the poinsettia varieties 'Vintage', 'Christmas Feelings' and 'Christmas Feelings Pearl' showed the expected allele configuration: homozygous wild-type, heterozygous and homozygous mutations, respectively (Figure 2). The mutated allele was detected with high accuracy at dilutions of 1:1, 5:1 and 10:1 in wild-type DNA, whereas the approach was not sensitive for detecting the mutations at dilutions of 100:1 and 500:1 (Figure 2). Although IRD labelling was highly sensitive with low background noise, only faint amplification could be observed for the 20:1 and 50:1 dilutions of mutated alleles in wild-type DNA, but the same faint signal could be detected in DNA of the 'Vintage' variety (RR) indicating background noise due to slippage of the DNA polymerase. Therefore, a reliable result could not be drawn for those samples. Nonetheless, the fluorescent labelling of PCR fragments approach followed by

TABLE 1 List of primers used in this study

Primer name	Sequence	Analysis
*Bract1*_F	GTAAAACGACGGCCAGTTGGCCTGCCTTTTAGAGAAA	Genotyping in acrylamide gel
*Bract1*_R	ACAAGTTCAGGGGGCTGAG	
IRD_{700}-labelled M13	FAM-GTAAAACGACGGCCAGT	
*Bract1*_F	GTAAAACGACGGCCAGTTGGCCTGCCTTTTAGAGAAA	First PCR round amplicon sequencing *Poinsettia samples*
*Bract1*_R	CAGGAAACAGCTATGACACAAGTTCAGGGGGCTGAG	
*Bract1*_F	GTAAAACGACGGCCAGTTGGCCTGCCTTTTAGAGAAA	First PCR round amplicon sequencing *Plasmid samples*
*Bract1*_plasmid_R	CAGGAAACAGCTATGACGGCCCTCTTTCTTCATATTTTG	
Tag1_F	AAGTCTTCGTGTAAAACGACGGCCAGT	Second PCR round for amplicon sequencing
Tag1_R	AATGCGCTATCAGGAAACAGCTATGAC	
Tag2_F	ACCTCATCTTGTAAAACGACGGCCAGT	
Tag2_R	ACGTGTTACTCAGGAAACAGCTATGAC	
Tag3_F	AACAACTGCTGTAAAACGACGGCCAGT	
Tag3_R	AAGCTCACTTCAGGAAACAGCTATGAC	
Tag4_F	ACAGTTAGCTGTAAAACGACGGCCAGT	
Tag4_R	ACGATACGTTCAGGAAACAGCTATGAC	
Tag5_F	ACAACCAGTTGTAAAACGACGGCCAGT	
Tag5_R	ACATGAGGTTCAGGAAACAGCTATGAC	
Tag6_F	AATCCTGGATGTAAAACGACGGCCAGT	
Tag6_R	AGAGTTGCTTCAGGAAACAGCTATGAC	
Tag7_F	ATGTTAGGATGTAAAACGACGGCCAGT	
Tag7_R	AAGACCGAATCAGGAAACAGCTATGAC	
Tag8_F	AGTATTCGATGTAAAACGACGGCCAGT	
Tag8_R	ATCGGACCATCAGGAAACAGCTATGAC	
Tag9_F	AAGCTACCATGTAAAACGACGGCCAGT	
Tag9_R	AAGGTCCTATCAGGAAACAGCTATGAC	
Tag10_F	ACTTGACGATGTAAAACGACGGCCAGT	
Tag10_R	ATCATTCGATCAGGAAACAGCTATGAC	

Abbreviation: PCR, polymerase chain reaction.

electrophoresis resolution could reliably detect heterozygous mutations in poinsettia DNA samples up to a 10:1 dilution.

3.2 | Amplicon-seq approach to identify heterozygous mutations

Our amplicon-seq analysis was performed in two experiments: (i) plasmid sequencing, where we used the plasmids *35S::Bract1* and *35S::Bract1_mut* to first evaluate if the approach would be suitable to detect the *Bract1* mutation and to determine the intrinsic error rates, and (ii) poinsettia sequencing, where we sequenced the *Bract1* locus from PCR products amplified from genomic DNA obtained from the varieties Vintage, Christmas Feelings and Christmas Feelings Pearl, and the serial dilutions 1:1, 5:1, 10:1, 20:1, 50:1, 100:1 and 500:1 were employed. The sequencing strategy consisted of a first round of PCR with *Bract1*-specific primers tagged with M13 tails followed by a second PCR with complementary M13 primers including tags to differentiate PCR fragments arising from each variety and PCR replicates (Figure 1).

Three independent PCRs were performed for each plasmid for the first amplicon-seq experiment, whereas for the second experiment (the poinsettia sample and serial dilutions), five independent PCRs were performed for each sample. The tags used for each sample are available in Table 1. The amplicon-seq experiments showed an average sequencing coverage of 135,890 reads for the plasmids and an average coverage of sequencing of 62,995 for the second experiment (Figure 3). The mutated *Bract1* allele (containing a 4-bp deletion) was detected in all analysed samples, including the *35S::Bract1* plasmid and the homozygous dominant variety Vintage (Data S1). On the other hand, the wild-type *Bract1* allele was also detected in the *35S::Bract1_mut* plasmid and the homozygous recessive variety Christmas Feelings Pearl. Figure 4 shows the percentage of the *Bract1*-mutated allele detected in each of the varieties and dilutions. The heterozygous variety Christmas Feelings showed a ratio of 1:1 from both alleles. The serial dilutions showed a serial

TABLE 2 Tag distribution for the amplicon-sequencing experiments

Experiment	Sample	Forward tag	Reverse tag (one tag per PCR replicate)
Plasmid sequencing	*35S::Bract1*	Tag_F1	Tag_R1-R3
	35S::Bract1_mut	Tag_F2	Tag_R4-R6
Poinsettia sequencing	Vintage	Tag_F1	Tag_R1-R5
	Chr. Feelings	Tag_F2	Tag_R1-R5
	Chr. Feelings Pearl	Tag_F3	Tag_R1-R5
	1:1	Tag_F4	Tag_R1-R5
	5:1	Tag_F5	Tag_R1-R5
	10:1	Tag_F6	Tag_R6-R10
	20:1	Tag_F7	Tag_R6-R10
	50:1	Tag_F8	Tag_R6-R10
	100:1	Tag_F9	Tag_R6-R10
	500:1	Tag_F10	Tag_R6-R10

Note: Sequence information for each of the tags is available in Table 1.
Abbreviation: PCR, polymerase chain reaction.

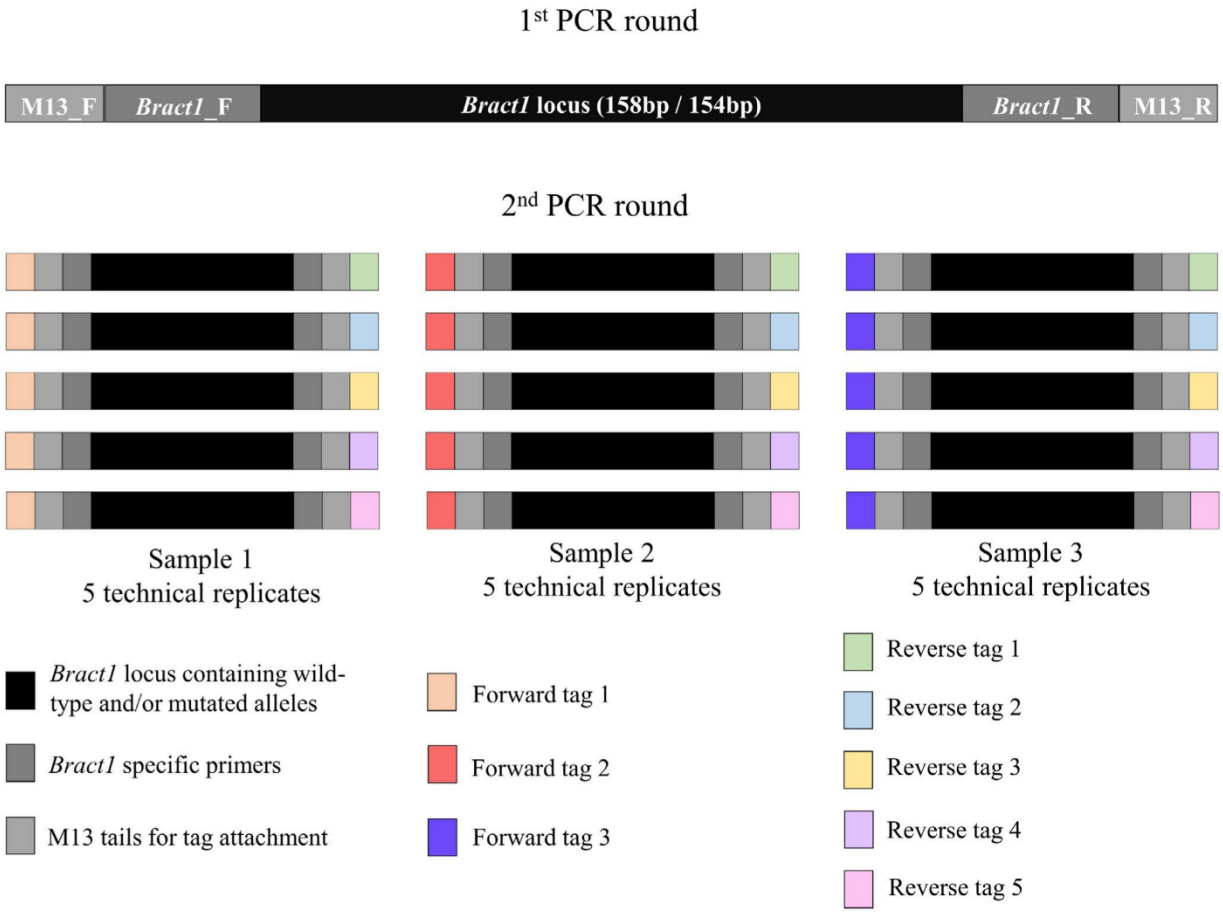


FIGURE 1 Schematic overview of the amplicon-sequencing (amplicon-seq) strategy for the *Bract1* locus. The first polymerase chain reaction (PCR) round amplifies the *Bract1* locus containing the 4-bp indel. The second PCR round introduces tags for each sample and PCR replicate prior to sequencing. The picture depicts an example for three hypothetical samples. The tag distribution for each sample can be seen in Table 2 [Colour figure can be viewed at wileyonlinelibrary.com]

reduction in the detection of the mutated allele, ranging from 36.1% (1:1) to 0.3% (500:1).

3.3 | Detection accuracy of heterozygous mutations

To address the detection accuracy of the amplicon-seq assay for the detection of heterozygous mutations in the poinsettia breeding pipeline, we compared the serial dilutions containing the lower concentrations of mutated alleles (20:1, 50:1, 100:1 and 500:1) to the control samples (*35S::Bract1* and Vintage). As previously mentioned, the mutated allele was detected in the control samples, which may be explained by the fact that library construction for sequencing relies on PCR, which is prone to polymerase slippage at repeat motifs, thereby leading to minor amplicon products that differ by multiples of the length of the repeat unit (also known as stutters; Ellegren, 2004). Therefore, the control samples used a threshold for the limit of detection of mutated alleles in our assay.

When compared with the *35S::Bract1* plasmid control, the dilutions 20:1 and 50:1 showed significant differences (Kruskal–Wallis,

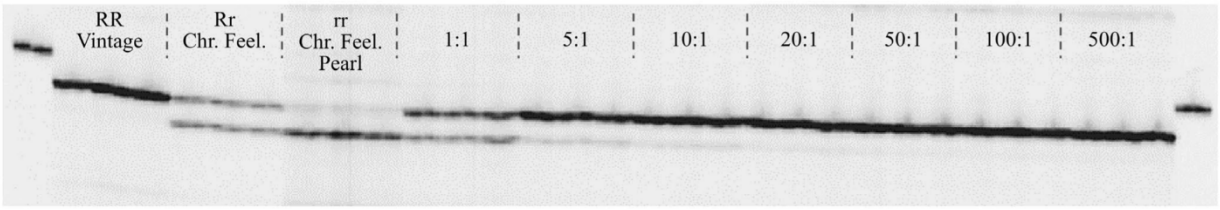


FIGURE 2 Detection of *Bract1* alleles in poinsettia genotypes and in serial dilutions of wild-type (R) and mutated (r) *Bract1* alleles using fluorescent labelling with polymerase chain reaction (PCR) fragments. For each poinsettia variety and dilution, three independent PCR analyses were performed. A size marker (~200 bp) is shown on both sides of the gel image. Chr. Feel., 'Christmas Feelings' variety; Chr. Feel. Pearl, 'Christmas Feelings Pearl' variety

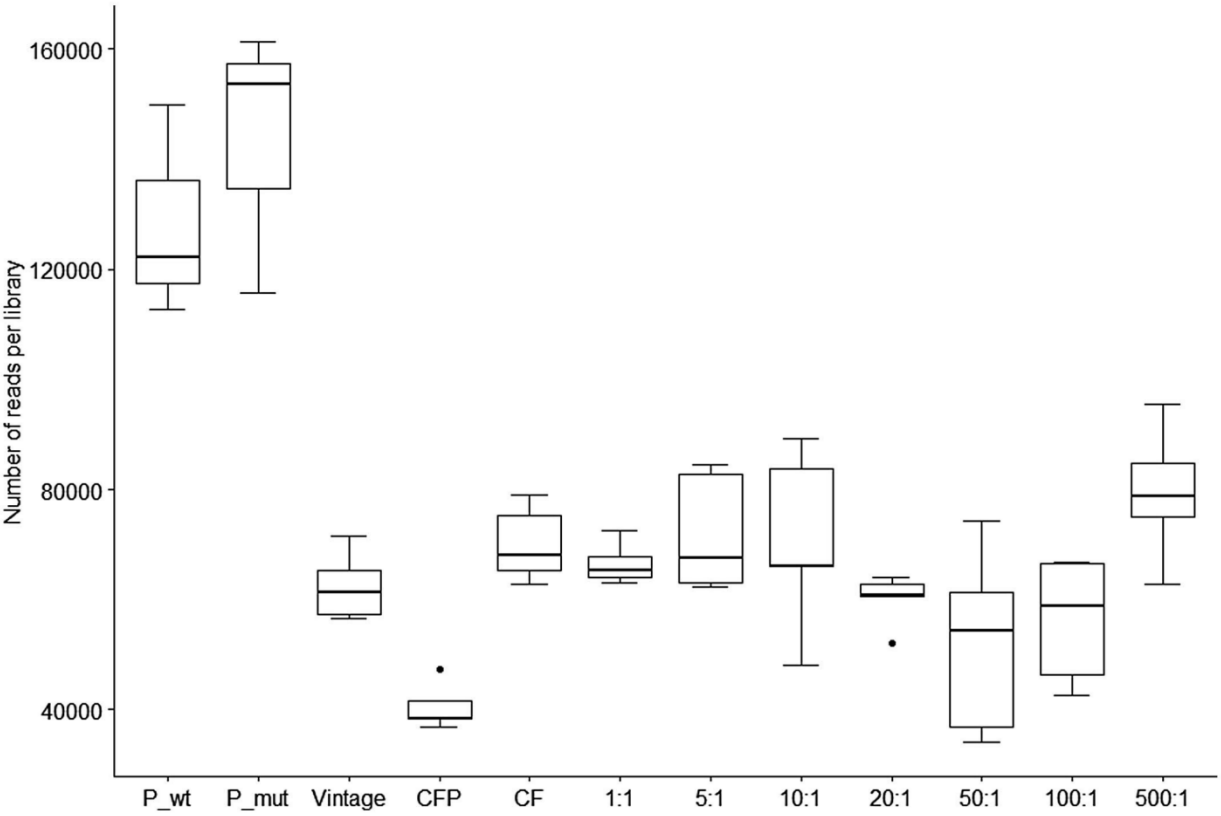


FIGURE 3 Sequencing coverage for two amplicon-sequencing (amplicon-seq) experiments. Boxplots for the plasmids *35S::Bract1* (P_wt) and *35S::Bract1_mut* (P_mut) were calculated with three independent replicates. Boxplots for the poinsettia samples and serial dilutions were calculated with five independent replicates. CF, 'Christmas Feelings' variety; CFP, 'Christmas Feelings Pearl' variety

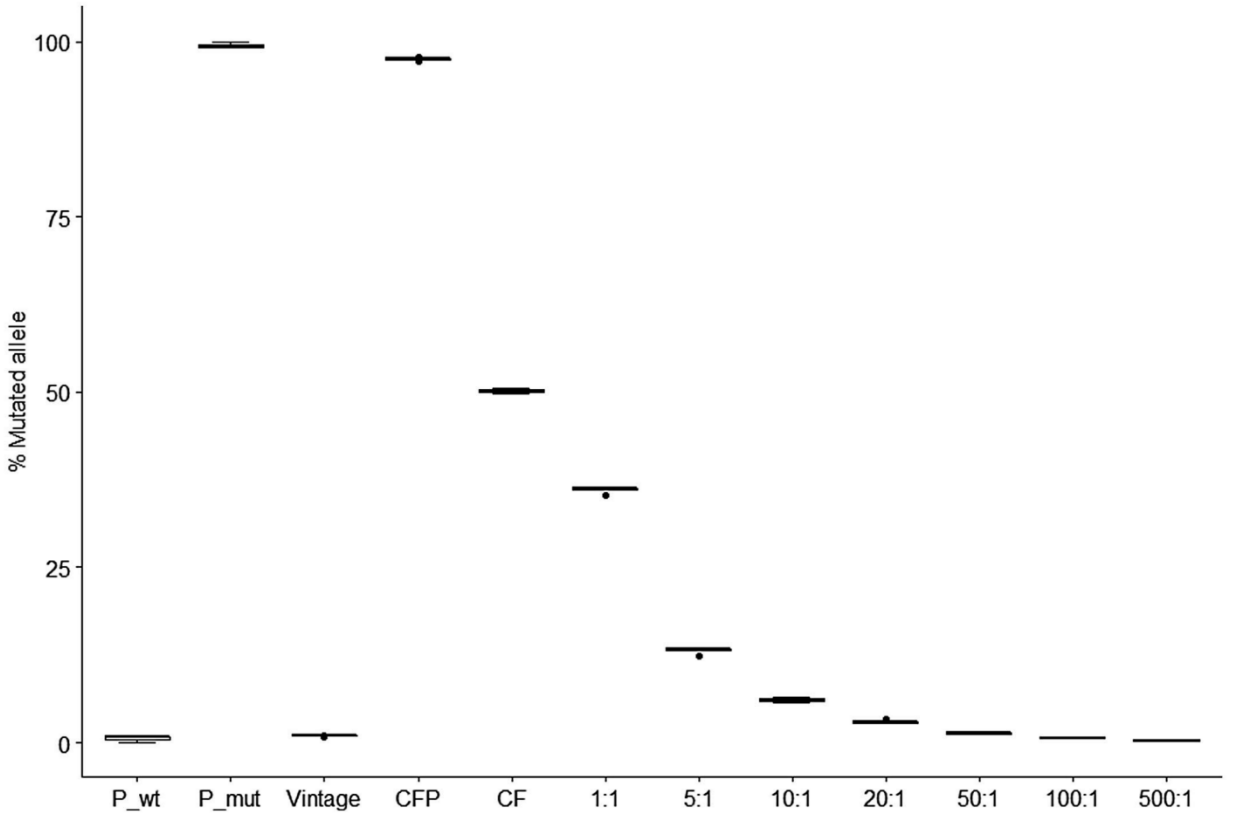


FIGURE 4 Detection of the *Bract1* mutated allele in amplicon-sequencing (amplicon-seq) experiments. Boxplots for the plasmids *35S::Bract1* (P_wt) and *35S::Bract1_mut* (P_mut) were calculated with three independent replicates. Boxplots for the poinsettia samples and serial dilutions were calculated with five independent replicates. CF, 'Christmas Feelings' variety; CFP, 'Christmas Feelings Pearl' variety

$p \leq .05$) for the detection level of the mutated allele, whereas the dilution 100:1 showed no differences, and the 500:1 dilution fell under the threshold of the assay (Figure 5a). The dilutions 20:1 and 50:1 were also significantly different (Kruskal–Wallis, $p \leq .05$) from the 'Vintage' variety, whereas the dilutions 100:1 and 500:1 fell under the detection threshold (Figure 5b). These results show that the amplicon-seq approach can accurately detect the presence of the *Bract1* mutation at dilutions of at least 50:1 (wild-type allele:mutant allele).

4 | DISCUSSION

Ionizing radiation is an important and frequently used tool for mutation breeding for bract colour variants in poinsettia. In contrast to ethyl methanesulphonate (EMS) chemical mutagenesis, which produces point mutations with high frequency (Greene et al., 2003), ionizing radiation (e.g., X-rays and γ-rays) induces DNA oxidative damage, such as double-strand breaks (DSBs), base substitutions, deletions and chromosomal alterations, at a lower frequency, thereby commonly resulting in the loss of gene function (Jo & Kim, 2019; Kazama et al., 2011; Morita et al., 2009). We have previously identified a short 4-bp repeat in the poinsettia anthocyanin-related *GST* gene (name *Bract1*). Upon X-ray treatment, progenies of irradiated plants show a deletion of one repeat (4 bp) in the *Bract1* locus at very high frequencies, leading to a frame shift in the coding region and a loss of function of the *Bract1* gene. This mutation is associated with the appearance of a white phenotype (Vilperte et al., 2020).

SSRs and short repeats are among the most variable and highly dynamic types of repetitive sequences in the genome (Ellegren, 2004; Verstrepen et al., 2005). These sequences are defined as short tandemly repeated DNA sequences that involve a repetitive unit of 1–9 bp and lengths of up to 100 nucleotides (nt; Fan & Chu, 2007; Gemayel et al., 2010). SSR regions are highly unstable, where mutations occur in the form of the addition or deletion of repeat units at considerably higher rates than in other parts of the genome (10 to 100,000 times higher; Gemayel et al., 2010; Schlötterer, 2000). The most evident source of SSR diversity is the slippage of DNA polymerases during SSR replication (Baptiste & Eckert, 2012; Brandström et al., 2008; Webster & Hagberg, 2007). Such slippage also occurs in in vitro SSR PCR amplification, which results in erroneous SSR alleles, thereby making accurate SSR genotyping challenging. In addition, gel electrophoresis, the most commonly used approach to detecting SSR PCR products, does not enable high-throughput applications and is not accurate due to its low resolution (Li et al., 2017). Although we were able to apply our gel electrophoresis approach to

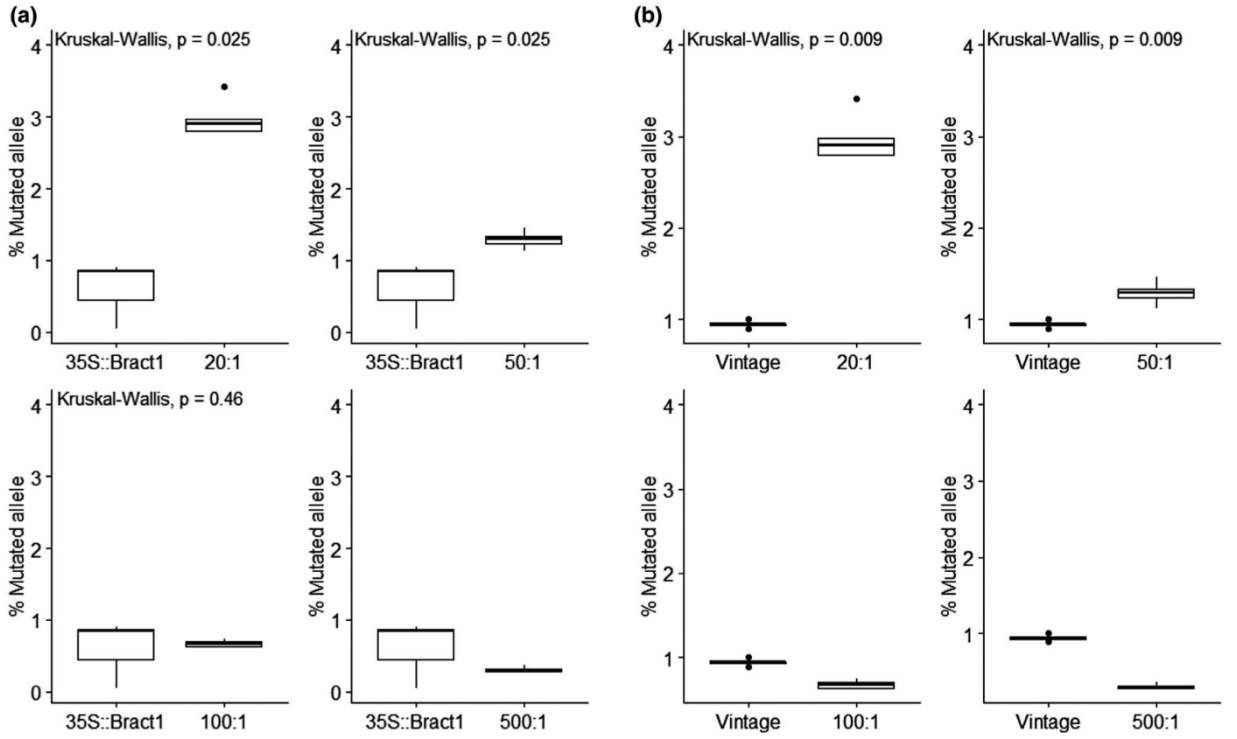


FIGURE 5 Limit of detection of heterozygous mutations using amplicon sequencing. (a) Comparison of amplicon-sequencing coverage of the mutant allele in serial dilutions with the *35S::Bract1* plasmid. (b) Comparison of amplicon-sequencing coverage of the mutant allele in serial dilutions of DNA from the homozygous dominant variety 'Vintage' mixed with DNA from the white variety 'Christmas Feelings Pearl'. Boxplots for the plasmids *35S::Bract1* and *35S::Bract1_mut* were calculated with three independent replicates. Boxplots for the poinsettia samples and serial dilutions were calculated with five independent replicates. The Kruskal–Wallis test for non-parametric data ($p \leq .05$) for pairwise comparisons between samples was performed using in-house R scripts

detect SSR *Bract1* mutations, it was not sensitive enough to detect heterozygous mutations at low frequencies. Moreover, only a small number of samples could be screened at a time.

Amplicon-seq is an increasingly frequently utilized genotyping approach that provides a cost-effective strategy; that is, it is time efficient and less laborious than conventional electrophoresis-based approaches to profile short targeted gene sequences (Early et al., 2019; Li et al., 2019). High-throughput approaches based on next-generation sequencing (NGS) have only rarely been applied to SSR genotyping but have nonetheless yielded helpful outputs. Yang et al., (2019) applied an approach called target SSR-seq, which combines the multiplexed amplification of SSRs with high-throughput sequencing and can be used for genetic background selection, gene and QTL mapping, and molecular breeding of cucumber varieties. A similar approach (AmpSeq-SSR) was used for fingerprinting rice varieties, as well as for mapping resistance genes (Li et al., 2017). Our amplicon-seq approach is similar to both Target SSR-seq and AmpSeq-SSR, and it was determined to be reliable and accurate in detecting low-frequency heterozygous mutations in poinsettia.

Despite the importance of SSR genotyping by NGS and its recent applications in many fields of molecular biology, employing this method, especially for small insertions and deletions (*indels*) analysis, remains challenging and error-prone (Albers et al., 2011; Treangen & Salzberg, 2011; Wang et al., 2011). In our amplicon-seq experiments, both negative controls for the *Bract1* mutated allele (*35S::Bract1* plasmid and 'Vintage' variety) and the positive controls (*35S::Bract1* and 'Christmas Feelings Pearl') showed a low background of mutated products, most likely the results of slippage that occurred during PCR. In addition to error introduced by the slippage of in vitro SSR PCR, there is an increased sequencing error in such regions due to imprecise base calling of NGS platforms (Albers et al., 2011). Studies using plasmids containing artificially inserted tandem repeat sequences have exhibited stutters due to slippage, which is strongly correlated with the number of repeat units and repeat unit length (Fungtammasan et al., 2015; Shinde, 2003). It is therefore possible that the amplicon variants observed in our study are likely attributable to polymerase slippage and/or base calling errors.

It is important to mention that the role of the *Bract1* mutation in the appearance of white phenotypes in poinsettia is not only relevant for the varieties analysed in this study. Poinsettia varieties with red bracts homozygous for the wild-type allele of *Bract1* differ for several other characteristics, for example, bract and shoot morphology, and production parameters. As poinsettia is a vegetatively propagated crop, introgression of the mutated allele of *Bract1* into elite varieties is not an option leaving de novo mutagenesis as the only option for breeding white variants of existing varieties. The high mutagenicity

of *Bract1* therefore provides the opportunity to obtain such mutants from commercially important homozygous red varieties. However, mutations, either natural or induced, can occur in several genes involved in the anthocyanin biosynthesis/accumulation and, therefore, lead to acyanic phenotypes. An insertion in a *leucoanthocyanidin dioxygenase* (*LDOX*) gene led to the appearance of pomegranates containing a white phenotype when the allele is a recessive state (Ben-Simhon et al., 2015). A frameshift mutation in a *dihydroflavonol 4-reductase* (*DFR*) gene is responsible for a rare white flower phenotype in *Mimulus lewisii* (Wu et al., 2013). The insertion of a retrotransposable element into the *F3H* 5′-upstream region caused a reduction in transcript expression and thus responsible for the white petal phenotype in *Torenia fournieri* (Nishihara et al., 2014). Analysing possible mutations in other anthocyanin-related genes in a wider range of varieties (e.g., from different breeding programmes) would generate valuable information on alternative targets that could be used in colour breeding in poinsettia.

Nevertheless, we were able to distinguish the background noise from true mutations down to a dilution of 50:1. This finding enables us to improve the generation of novel colour mutants in poinsettias in two important aspects. The first is that screens in homozygous red varieties may now become more feasible and considerably more effective compared with previous screens, as heterozygotes can be readily detected and used in a second round of mutagenic treatment, leading to white varieties. The second aspect is that considerably lower dosages of ionizing radiation may now be employed for the generation of heterozygotes from homozygotes, thereby reducing the mutational load that is associated with the application of ionizing radiation. In the long term, this approach avoids the accumulation of negative mutations in the poinsettia breeding pool.

5 | CONCLUSIONS

The amplicon-seq approach utilized in our study exhibited higher accuracy and sensitivity to detect heterozygous mutations in poinsettia DNA samples. In addition, this approach demonstrated the ability to detect low-frequency mutations at a sensitivity of up to one mutant allele in 50 nonmutated alleles, thereby increasing the possibility of high-throughput screenings. Because the current approach that is applied for mutagenic breeding in poinsettia is based on the use of heterozygous plants and the subsequent selection of white sports, the development of a strategy to detect heterozygous mutations arising from radiation mutagenesis of homozygous plants is a valuable tool to implement in the breeding process. This approach may enable an expansion of the genetic resources available for the development of new varieties, especially those related to colour traits, as well as enabling a reduction in the mutation load in poinsettia gene pools.

ACKNOWLEDGEMENTS

We would like to thank Guido von Tubeuf for providing the poinsettia varieties. This project has received funding from the European Union's Horizon 2020 Framework Programme (research and innovation programme) under the Marie Skłodowska Curie Grant Agreement No. 675657 Flower Power. This funding body had no role in the design of the study, in the 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.

CONFLICT OF INTERESTS

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.

AUTHOR CONTRIBUTIONS

V. V., T. D. and R. B. conceived the research. V. V. and T. D. designed the experiments. R. B. provided the plant material. V. V. conducted the laboratory and bioinformatics analyses. V. V. and T. D. wrote the manuscript. All authors revised and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The DNA sequence from the *Bract1* gene is available in the GenBank repository under Accession Number MW718861. The CDS sequences from several poinsettia genotypes are available in the GenBank repository under Accession Numbers MW718847–MW718860. The NGS data generated in this study are available on request from the authors.

ORCID

Thomas Debener https://orcid.org/0000-0002-2012-2246

REFERENCES

Albers, C. A., Lunter, G., MacArthur, D. G., McVean, G., Ouwehand, W. H., & Durbin, R. (2011). Dindel: Accurate indel calls from short-read data. *Genome Research*, *21*(6), 961–973. https://doi.org/10.1101/gr.112326.110

Baptiste, B. A., & Eckert, K. A. (2012). DNA polymerase kappa microsatellite synthesis: Two distinct mechanisms of slippage-mediated errors. *Environmental and Molecular Mutagenesis*, *53*(9), 787–796. https://doi.org/10.1002/em.21721

Ben-Simhon, Z., Judeinstein, S., Trainin, T., Harel-Beja, R., Bar-Ya'akov, I., Borochov-Neori, H., & Holland, D. (2015). A "white" anthocyanin-less pomegranate (*Punica granatum* L.) caused by an insertion in the coding region of the leucoanthocyanidin dioxygenase (*LDOX*; *ANS*) gene. *PLoS One*, *10*(11), e0142777. https://doi.org/10.1371/journal.pone.0142777

Brandström, M., Bagshaw, A. T., Gemmell, N. J., & Ellegren, H. (2008). The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Molecular Biology and Evolution*, *25*(12), 2579–2587. https://doi.org/10.1093/molbev/msn201

Datta, S. K. (2020). Induced mutations: Technological advancement for development of new ornamental varieties. *The Nucleus*, *63*(2), 119–129. https://doi.org/10.1007/s13237-020-00310-7

Early, A. M., Daniels, R. F., Farrell, T. M., Grimsby, J., Volkman, S. K., Wirth, D. F., MacInnis, B. L., & Neafsey, D. E. (2019). Detection of low-density *Plasmodium falciparum* infections using amplicon deep sequencing. *Malaria Journal*, *18*(1), 219. https://doi.org/10.1186/s12936-019-2856-1

Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Fan, H., & Chu, J.-Y. (2007). A brief review of short tandem repeat mutation. *Genomics, Proteomics & Bioinformatics*, *5*(1), 7–14. https://doi.org/10.1016/S1672-0229(07)60009-6

FAO, F. and A. O. of the U. N., & IAEA, I. A. E. A. (2019). IAEA Mutant Varieties Database [Text]. IAEA. https://mvd.iaea.org/

Fungtammasan, A., Ananda, G., Hile, S. E., Su, M.-S.-W., Sun, C., Harris, R., Medvedev, P., Eckert, K., & Makova, K. D. (2015). Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Research*, *25*(5), 736–749. https://doi.org/10.1101/gr.185892.114

Gemayel, R., Vinces, M. D., Legendre, M., & Verstrepen, K. J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual Review of Genetics*, *44*(1), 445–477. https://doi.org/10.1146/annurev-genet-072610-155046

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., Enns, L. C., Burtner, C., Johnson, J. E., Odden, A. R., Comai, L., & Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, *164*(2), 731–740.

Gu, Z., Chen, H., Yang, R., & Ran, M. (2018). Identification of DFR as a promoter of anthocyanin accumulation in poinsettia (*Euphorbia pulcherrima*, Willd. Ex Klotzsch) bracts under short-day conditions. *Scientia Horticulturae*, *236*, 158–165. https://doi.org/10.1016/j.scienta.2018.03.032

Ibrahim, R., Ahmad, Z., Salleh, S., Hassan, A. A., & Ariffin, S. (2018). Mutation breeding in ornamentals. In J. Van Huylenbroeck (Ed.), *Ornamental crops* (pp. 175–211). Springer International Publishing. https://doi.org/10.1007/978-3-319-90698-0_8

Jo, Y. D., & Kim, J.-B. (2019). Frequency and spectrum of radiation-induced mutations revealed by whole-genome sequencing analyses of plants. *Quantum Beam Science*, *3*(2), 7. https://doi.org/10.3390/qubs3020007

Kazama, Y., Hirano, T., Saito, H., Liu, Y., Ohbu, S., Hayashi, Y., & Abe, T. (2011). Characterization of highly efficient heavy-ion mutagenesis in *Arabidopsis thaliana*. *BMC Plant Biology*, *11*, 161. https://doi.org/10.1186/1471-2229-11-161

Li, L., Fang, Z., Zhou, J., Chen, H., Hu, Z., Gao, L., Chen, L., Ren, S., Ma, H., Lu, L., Zhang, W., & Peng, H. (2017). An accurate and efficient method for large-scale SSR genotyping and applications. *Nucleic Acids Research*, *45*(10), e88. https://doi.org/10.1093/nar/gkx093

Li, T., Fang, Z., Peng, H., Zhou, J., Liu, P., Wang, Y., Zhu, W., Li, L., Zhang, Q., Chen, L., Li, L., Liu, Z., Zhang, W., Zhai, W., Lu, L., & Gao, L. (2019). Application of high-throughput amplicon sequencing-based SSR genotyping in genetic background screening. *BMC Genomics*, *20*(1), 444. https://doi.org/10.1186/s12864-019-5800-4

Morita, R., Kusaba, M., Iida, S., Yamaguchi, H., Nishio, T., & Nishimura, M. (2009). Molecular characterization of mutations induced by gamma irradiation in rice. *Genes & Genetic Systems*, *84*(5), 361–370. https://doi.org/10.1266/ggs.84.361

Moustaka, J., Panteris, E., Adamakis, I.-D.- S., Tanou, G., Giannakoula, A., Eleftheriou, E. P., & Moustakas, M. (2018). High anthocyanin accumulation in poinsettia leaves is accompanied by thylakoid membrane unstacking, acting as a photoprotective mechanism, to prevent ROS formation. *Environmental and Experimental Botany*, *154*, 44–55. https://doi.org/10.1016/j.envexpbot.2018.01.006

Nielsen, M. D., Farestveit, B., & Andersen, A. S. (2003). Adventitious shoot development from decapitated plants of peri-clinal chimeric poinsettia plants (*Euphorbia pulcherrima* Willd ex Klotsch). *European Journal of Horticultural Science*, *68*(4), 161–168.

Nishihara, M., Yamada, E., Saito, M., Fujita, K., Takahashi, H., & Nakatsuka, T. (2014). Molecular characterization of mutations in white-flowered torenia plants. *BMC Plant Biology*, *14*, 86. https://doi.org/10.1186/1471-2229-14-86

Nitarska, D., Stefanini, C., Haselmair-Gosch, C., Miosic, S., Walliser, B., Mikulic-Petkovsek, M., Regos, I., Slatnar, A., Debener, T., Terefe-Ayana, D., Vilperte, V., Hadersdorfer, J., Stich, K., & Halbwirth, H. (2018). 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 Biology*, *18*(1), 216.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, *109*(6), 365–371. https://doi.org/10.1007/s004120000089

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, *18*(2), 233–234. https://doi.org/10.1038/72708

Schum, A., & Preil, W. (1998). Induced mutations in ornamental plants. In S. M. Jain, D. S. Brar, & B. S. Ahloowalia (Eds.), *Somaclonal variation and induced mutations in crop improvement* (pp. 333–366). Springer. https://doi.org/10.1007/978-94-015-9125-6_17

Shinde, D. (2003). *Taq* DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: $(CA/GT)_n$ and $(A/T)_n$ microsatellites. *Nucleic Acids Research*, *31*(3), 974–980. https://doi.org/10.1093/nar/gkg178

Slatnar, A., Mikulic-Petkovsek, M., Veberic, R., Stampar, F., & Schmitzer, V. (2013). Anthocyanin and chlorophyll content during poinsettia bract development. *Scientia Horticulturae*, *150*, 142–145. https://doi.org/10.1016/j.scienta.2012.10.014

Stewart, R. N., Asen, S., Massie, D. R., & Norris, K. H. (1980). The anthocyanin and flavonol composition of three families of poinsettia color sports. *Journal of Heredity*, *71*(2), 140–142. https://doi.org/10.1093/oxfordjournals.jhered.a109329

Treangen, T. J., & Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics*, *13*(1), 36–46. https://doi.org/10.1038/nrg3117

Trejo, L., Feria Arroyo, T. P., Olsen, K. M., Eguiarte, L. E., Arroyo, B., Gruhn, J. A., & Olson, M. E. (2012). Poinsettia's wild ancestor in the Mexican dry tropics: Historical, genetic, and environmental evidence. *American Journal of Botany*, *99*(7), 1146–1157

van der Ploeg, R. (2018). FloraCulture International. International Association of Horticultural Producers (IAHP). https://www.floraculture.eu/digital-magazine/fci-november-december-2018/

Verstrepen, K. J., Jansen, A., Lewitter, F., & Fink, G. R. (2005). Intragenic tandem repeats generate functional variability. *Nature Genetics*, *37*(9), 986–990. https://doi.org/10.1038/ng1618

Vilperte, V., Boehm, R., & Debener T. (2021). A highly mutable GST is essential for bract colouration in *Euphorbia pulcherrima* Willd. Ex Klotsch. *BMC Genomics*, *22*(1). https://doi.org/10.1186/s12864-021-07527-z

Vilperte, V., Lucaciu, C. R., Halbwirth, H., Boehm, R., Rattei, T., & Debener, T. (2019). Hybrid de novo transcriptome assembly of poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) bracts. *BMC Genomics*, *20*(1). https://doi.org/10.1186/s12864-019-6247-3

Wang, D. Y., Chang, C.-W., Lagacé, R. E., Oldroyd, N. J., & Hennessy, L. K. (2011). Development and validation of the AmpF*ℓ*STR® Identifiler® Direct PCR Amplification Kit: A multiplex assay for the direct amplification of single-source samples. *Journal of Forensic Sciences*, *56*(4), 835–845. https://doi.org/10.1111/j.1556-4029.2011.01757.x

Webster, M. T., & Hagberg, J. (2007). Is there evidence for convergent evolution around human microsatellites? *Molecular Biology and Evolution*, *24*(5), 1097–1100. https://doi.org/10.1093/molbev/msm051

Wu, C. A., Streisfeld, M. A., Nutter, L. I., & Cross, K. A. (2013). The genetic basis of a rare flower color polymorphism in *Mimulus lewisii* provides insight into the repeatability of evolution. *PLoS One*, *8*(12), e81173. https://doi.org/10.1371/journal.pone.0081173

Yang, J., Zhang, J., Han, R., Zhang, F., Mao, A., Luo, J., Dong, B., Liu, H., Tang, H., Zhang, J., & Wen, C. (2019). Target SSR-Seq: A novel SSR genotyping technology associate with perfect SSRs in genetic

analysis of cucumber varieties. *Frontiers in Plant Science*, *10*, https://doi.org/10.3389/fpls.2019.00531

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Vilperte V, Boehm R, Debener T. Development of a multiplex amplicon-sequencing assay to detect low-frequency mutations in poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) breeding programmes. *Plant Breed*. 2021;140:497–507. https://doi.org/10.1111/pbr.12925

4.1 Supporting information

Additional File S1. Number of reads containing either the Bract1 wild-type or mutated alleles for each replicate of each sequencing library (Available as electronic appendix).

5 General discussion

The main goal of this study was to unravel the molecular mechanism involved in the appearance of high frequency acyanic mutants after mutagenesis treatment of red poinsettia genotypes. First, a *de novo* transcriptome for poinsettia bracts was generated and used as the reference for expression analysis (manuscript 1). Secondly, a thorough molecular characterization of a candidate gene involved in the anthocyanin pathway was performed (manuscript 2). Lastly, a multiplex assay was developed to identify low-frequency mutations and to improve poinsettia breeding programs. The main results that were obtained in the three studies presented in this thesis are already discussed in the respective manuscripts. More general aspects are discussed in this chapter to connect the findings of the different manuscripts and to give an outlook on the future use of the generated information.

5.1 The transcriptome of poinsettia bracts

Since the start of its active breeding efforts in the 1950s, red poinsettias are still the dominant varieties in breeding programs. Nonetheless, a range of other colours, especially white varieties, are becoming well accepted in the current market (van der Ploeg, 2018) and alternative colours may be economically interesting to specific niches. However, the efforts to understand the molecular mechanisms underlying the colour formation in poinsettia are still scarce. With that in mind, the first goal of this study was to generate valuable genetic data for the species and to utilize it towards the understanding of colour formation. A hybrid transcriptome assembly approach, which integrates short reads (Illumina) with long reads (PacBio), was used to create a poinsettia transcriptome and, therefore, serve as the foundation of this study. Such approaches have helped to improve transcriptome information for several plant species (Abdel-Ghany et al., 2016; Wang et al., 2016; Cheng et al., 2017; Hoang et al., 2017).

The final poinsettia bract transcriptome contains 288,524 contigs, with 33% showing confident protein annotation against the TAIR10 database. When working with non-model species, low gene annotation outcomes can be a common feature due to many reasons. The lack of a reference genome can hamper the transcriptome assembly, causing problems such as the sensitivity of alignment error due to paralogs and multigene families, production of artefactual chimaeras, problems reconstructing

transcript length, and potentially misestimating allelic diversity (Cahais et al., 2012; Palma-Silva et al., 2016; Ungaro et al., 2017). The quality of a transcriptome assembly can directly affect its functional annotation, which usually depends on the transfer of functional knowledge from one gene to another from different species, and several pitfalls in the process can lead to either incomplete or erroneous annotation of genes (Bolger et al., 2017).

Assessing the completeness of the transcriptome can also help to understand the overall quality of the transcriptome assembly. From a total of 1,440 BUSCO groups searched (Simão et al., 2015; Zdobnov et al., 2017), 77% of the poinsettia transcripts were categorized as complete, 10% as fragmented and 13% as missing BUSCOs. Although our transcriptome showed a lower level of completeness compared to other transcriptome studies (e.g. 91% in *Cinnamomum longepaniculatum* and 90% in *Noccaea caerulescens*) (Blande et al., 2017; Yan et al., 2017), the absence of a particular annotated gene or orthologue does not mean the lack of the gene-related function in the species. Functional annotation is highly dependent on complete gene models often transferred from model species (i.e. Arabidopsis); thus, for partial or incomplete gene models frequently obtained in *de novo* transcriptome assemblies, the bioinformatic tools are not always accurate to predict the correct function on a partially assembled gene (Bolger et al., 2017).

As the transcriptome is defined as the complete set of transcripts in a single cell and its quantities for a specific stage of development or physiological condition (Wang et al., 2009; Valdés et al., 2013), the choice of tissue used for a transcriptome assembly will directly impact its content. Varying levels of BUSCO completeness were observed when comparing different tissues of the same species (Babineau et al., 2017), thus indicating that tissue-specific transcripts account for differences in the expected gene space. The development of expression atlases (Libault et al., 2010; Fasoli et al., 2012; Kagale et al., 2016; Palovaara et al., 2017; Braich et al., 2019) are extremely helpful to obtain a nearly complete transcriptome picture and thus improve structural and functional gene annotation.

There are currently no consensus or optimal methodologies in place for creating and analysing the complexity of *de novo* transcriptome assemblies, especially for non-model species. Different species (with their different tissues, ploidy levels and genomic complexities), sequencing protocols, and parameter settings require different

approaches to achieve the most reliable results (Hölzer and Marz, 2019). Nonetheless, transcriptome studies are a powerful tool to reconstruct and quantify whole transcriptomes and to generate useful information for breeding purposes (Conesa et al., 2016; Hrdlickova et al., 2017).

5.2 Improving and expanding the knowledge on the genetic basis of bract colouration in poinsettia

The first attempts to decipher the inheritance of bract colour in poinsettia was done by Robert N. Stewart in 1960, whereby performing crosses between red and white genotypes, he concluded that white bracts differ from red bracts by a single recessive factor (*wh*) and that the red factor (*WH*) is completely dominant over white (Stewart, 1960). Over the years, this knowledge has been applied to poinsettia breeding programs, but not much effort has been made to identify the gene(s) responsible for the white phenotype. Therefore, the main goal of this thesis was to identify the causal gene and the genetic mechanisms involved in the appearance of white phenotypes.

In the first chapter of this thesis, we aimed to characterize possible gene expression differences in anthocyanin-related pathways between red and white poinsettia genotypes under similar genetic backgrounds (Chr. Feelings and Chr. Feelings Pearl) and during three bract developmental stages. As an outcome, we identified a *glutathione S-transferase* (*GST*) gene as being a putative candidate for further analysis since it showed a lower expression in the white variety, for both RNA-Seq and RT-qPCR analysis.

To further utilize the knowledge obtained in the first chapter of this thesis, the second chapter focused on thoroughly characterizing the expression, genetic variation and function of the poinsettia *GST* (named *Bract1*) as an anthocyanin transporter. We first observed a reduced transcript expression of the *Bract1* gene in six independent white mutants when compared to their red bracted counterpart. This corroborated with the results previously observed and enforced the hypothesis of the *Bract1* as a candidate gene. By sequencing the *Bract1* CDS from 14 red- and white-bracted poinsettia varieties, we identified two alleles in the gene: one present in all red varieties (here named the wild-type allele) and the second one, containing a 4 bp deletion located 8 bp upstream of the first exon-intron junction (here named the mutant allele), present only in the white varieties. The deletion is in a short repeat locus, resembling a short simple sequence repeat (SSR), with a tetranucleotide motif ($(CTTC)_3$) composition. This mutation leads to

a gain of a stop codon in the transcript and thus leading to a putative non-functional protein.

Although white poinsettia varieties are often obtained through radiation mutagenesis of the red ones, not all red varieties can produce white sports through radiation. Therefore, red poinsettia varieties are distinguished into ‘heterozygous’ and ‘homozygous’ for the colouration locus according to their ability to generate white sports and according to the segregation of red and white phenotypes in the progeny of crosses with white genotypes. By using a genotyping approach based on the fluorescent labelling of PCR fragments, we observed that previously categorized red heterozygous varieties (those that have been successfully used for mutagenesis) showed two distinct copies of the *Bract1* allele (wild-type and mutant alleles), while their white counterparts showed only the mutant one. On the other hand, homozygous red varieties (those previously unable to generate white sports) showed only the wild-type copy. These results indicate that the red bract colouration in poinsettia is a dominant trait.

In poinsettia breeding programmes, such homozygous red varieties are not used for mutagenesis breeding to obtain colour variants due to the lack of phenotypic traits for selection. However, they could be a valuable source of variation to overcome the narrow genetic base of modern breeding varieties in poinsettia (Trejo et al., 2012). With that in mind, we analysed the progeny of three homozygous irradiated varieties for putative *de novo* mutations in the *Bract1* gene. Mutant progeny (i.e. containing one copy of the mutated allele) was identified in all varieties, with a total of 7 mutants out of 756 screened progeny. These mutants would need to go to a second round of irradiation treatment to possibly generate white sports. Although such an approach would increase the length and costs of a breeding programme, harnessing the genetic variation of these varieties would be extremely beneficial in the long run. Nonetheless, one question remains open: Why do independent varieties submitted to irradiation treatment present the same type of mutation in the same locus?

Short repeat regions are extremely unstable, with mutation frequencies being much higher than in other parts of the genome (Schlötterer, 2000; Gemayel et al., 2010) and mainly occurring due to replication slippage (Fan and Chu, 2007; Gemayel et al., 2010). However, the use of ionizing radiation (e.g. X-rays and γ-rays) leads to DSBs, which are then repaired either via HR or NHEJ pathways, with the latter being predominant in plants (Puchta, 2005; Bleuyard et al., 2006; Roldán-Arjona and Ariza, 2009). Mutations in SSR

locations caused by DSB repair have been previously identified in plants (Pâques et al., 1998; Richard and Pâques, 2000; Verstrepen et al., 2005). SSR instability is shown to increases with plant development in Arabidopsis, and that NHEJ repair is more efficient in such cases (Golubov et al., 2010). Besides, abiotic stress has also been linked to SSR instability, with errors during NHEJ repair being most likely the instability cause (Yao and Kovalchuk, 2011). Mutations in SSR regions of EMS-treated sugarcane were linked to increased drought tolerance, but the authors could not elucidate if the SSR mutation was the potential cause (Khalil et al., 2018). Although our data cannot precisely answer the cause of such instability in the poinsettia *Bract1* gene, it indicates that particular short repeat sequences, similar to microsatellite sequences or so-called dynamic mutations, might be hot spots for genetic variability. Analysing other genomic short repeat regions in irradiate poinsettias would help to elucidate the molecular mechanisms behind such high-frequency mutations.

5.3 *Bract1* is a functional anthocyanin transporter in poinsettia bracts

Identifying the mutation in the *Bract1* gene as the most likely cause for the white phenotypes in poinsettia bracts, as well as elucidating the possible genetic mechanisms behind the mutation, was a great achievement for this project. However, as final proof of concept, we needed to confirm that *Bract1* is the functional anthocyanin transporter in poinsettia, as well as that the mutation leads to a loss of function. Therefore, we aimed to functionally characterize the *Bract1* by *in silico* and *in vivo* approaches: i) *Bract1* homology to known anthocyanin-related GSTs from other plant species and ii) *Bract1* ability to complement the colour phenotype in GST-deficient Arabidopsis mutants and (possibly) poinsettia bracts.

A phylogenetic tree containing 96 poinsettia-specific GST genes, together with seven anthocyanin-related GST from different species, revealed a high similarity of the *Bract1* gene with the other anthocyanin-related GSTs, with all of them belonging to the Phi class on the GST classification. Apart from *Bronze-2* from maize (Marrs et al., 1995), anthocyanin-related GSTs belong exclusively to the phi class, with the presence of two introns as a characteristic of this group of genes, such as *AN9* from petunia and *TT19* from Arabidopsis (Alfenito et al. 1998; Mueller et al., 2000), as well as *Bract1*. Furthermore, *Bract1* showed overall nucleotide and protein similarities with the other anthocyanin-related GSTs of 61.9% and 58.3%, respectively. Lastly, seven amino acid

residues, previously reported as specific to anthocyanin-related GSTs (Kitamura et al., 2012; Hu et al., 2016; Kou et al., 2019), were identified to be conserved in the *Bract1* gene and the anthocyanin-related GSTs: 2Val, 11Ala, 13Cys, 62Phe, 90Leu, 91Glu and 152Ser. All in all, the *in silico* analysis indicated that *Bract1* is the anthocyanin-related GST orthologue in poinsettia.

To examine the *in vivo* function of *Bract1* as an anthocyanin transporter, we tested the ability of *Bract1* cDNA to functionally complement the Arabidopsis GST mutant *tt19*, which is defective in the expression of anthocyanins in aboveground organs and seeds (Kitamura et al., 2004; Sun et al., 2012). Arabidopsis *tt19* mutants were transformed via the floral-dip method with distinct constructs harbouring the functional and the mutated *Bract1* alleles. Transgenic lines containing the functional *Bract1* allele displayed a purple hypocotyl phenotype at the seedling stage, while transgenic lines containing the mutated allele did not show complementation of the anthocyanin phenotype. Mutations in anthocyanin related GSTs leading to colourless phenotypes have been previously reported and thus corroborate with our results (Larsen et al., 2003; Cheng et al., 2015; Luo et al., 2018). These results not only emphasize the role of *Bract1* as an anthocyanin transporter in poinsettia but also demonstrate that a deletion in its coding region is not able to complement the anthocyanin phenotype.

The complementation of plant model mutants (e.g. the Arabidopsis *tt19* mutant) has been routinely applied in pigment-related functional studies (Hu et al., 2016; Pérez-Díaz et al., 2016; Liu et al., 2019), but final complementation in the real mutant will ultimately prove the gene function. Direct complementation of poinsettia white mutants with the functional *Bract1* was not possible during this thesis. Neither *Agrobacterium*-mediated infiltration nor biolistic particle delivery system (a.k.a. gene gun) transient assays were successful for the transformation of poinsettia bracts, even by testing different *Agrobacterium* strains (GV3101, GV2260, ABH, EHA105 and AGL1) and different protocol improvements (different rupture disk pressures and gold particle sizes) (data not shown).

Transient transformation in poinsettia has been successfully reported using leaf and bract-derived protoplasts, although only reporter genes were tested (Pitzschke and Persak, 2012). Attempts to stably transform poinsettia has been done using electrophoresis-based methods (Vik et al., 2001; Clarke et al., 2006), but no stable transgenic poinsettia was obtained. Successful stable transformation via *Agrobacterium*-mediated infiltration has been previously achieved, but the process is time-consuming

(Clarke et al., 2008; Islam et al., 2013). The construct used for the complementation of the Arabidopsis *tt19* mutant is currently being used for stable transformation of white bracted poinsettias, but no results were obtained so far since the experiments showed to be highly genotype- and explant-dependent, as well as time-consuming (data not shown). Altogether, our *in silico* and *in vivo* approaches show that *Bract1* is involved in anthocyanin transport in poinsettia bracts, as it shows high similarities with other anthocyanin-related GSTs and it was able to successfully complement the anthocyanin phenotype in Arabidopsis mutants.

It is important to mention that the role of the *Bract1* mutation in the appearance of white phenotypes in poinsettia may only be true for the varieties analysed in this study. Mutations, either natural or induced, can occur in several genes involved in the anthocyanin biosynthesis/accumulation and, therefore, lead to acyanic phenotypes. An insertion in a *leucoanthocyanidin dioxygenase* (*LDOX*) gene led to the appearance of pomegranates containing a white phenotype when the allele is in a recessive state (Ben-Simhon et al., 2015). A frameshift mutation in a *dihydroflavonol 4-reductase* (*DFR*) gene is responsible for a rare white flower phenotype in *Mimulus lewisii* (Wu et al., 2013). The insertion of a retrotransposable element into the *flavanone 3-hydroxylase* (*F3H*) 5′-upstream region caused a reduction in transcript expression and thus responsible for the white petal phenotype in *Torenia fournieri* (Nishihara et al., 2014). As the colour trait in poinsettia is known to be regulated by a dominant locus (Stewart, 1960), mutations in alternative gene targets could also occur. However, due to the genetic nature of the *Bract1* locus (short repeat region as mutation hotspot) and years of breeding selection, the identification of alternative mutations can be challenging. Nonetheless, analysing possible mutations in other anthocyanin-related genes in a wider range of varieties (e.g. from different breeding programs) would generate valuable information on alternative targets that could be used in colour breeding in poinsettia.

5.4 Development of an assay for mutation detection in poinsettia breeding programmes

Poinsettia mutation breeding is usually performed on cuttings that are irradiated with moderate dosages (~20-30 Gy) and mutants are phenotypically selected on side shoots of the originally irradiated plants. Still, as previously mentioned, heterozygous mutations arising from homozygous red genotypes occur at low frequencies and cannot be

phenotypically distinguished due to the dominance of *Bract1*. Now, with the identification of the *Bract1* mutation and the development of molecular markers that can identify heterozygous mutations, homozygous red varieties can be used in mutation breeding programmes. However, thousands of progenies need to be screened with molecular markers, thus making the process costly and time-consuming. With that in mind, this work aimed to develop a multiplex assay that could aid the detection of such low-frequency heterozygous mutations.

We created serial dilutions of wild-type (R) and mutated (r) *Bract1* alleles to simulate mutation frequencies and evaluate the potential for detecting mutations at such dilutions through two PCR-based techniques: i) the previously applied genotyping approach based on the fluorescent labelling of PCR fragments, and ii) Amplicon sequencing (Amp-Seq). The fluorescent labelling of PCR fragments approach could reliably detect heterozygous mutations in poinsettia DNA samples up to a 10:1 dilution. Although such a method can be easily applied in a routine laboratory set up, it might not sensitive enough to detect heterozygous mutations at low frequencies and only a small number of samples can be screened at a time. On the other hand, the Amp-Seq approach could accurately detect the presence of the *Bract1* mutation at dilutions of at least 50:1 copies. Moreover, such an approach enables the screening of multiple samples at the same time, thus increasing the throughput of the assay. In this experiment, we screened 50 different samples through the combinations of different primer tags. Adjusts can be made by using more tag combinations and/or by changing the sequencing coverage according to each experiment.

Although targeted sequencing-based methods are extremely helpful for breeding purposes (Jantzen et al., 2019; Nagano et al., 2020), the genotyping of short repeat sequences by such approaches remains challenging and error-prone. Errors introduced by the slippage of *in vitro* PCR are common, especially for repetitive regions, which is strongly correlated with the number of repeat units and repeat unit length of such regions (Shinde, 2003; Fungtammasan et al., 2015). Also, there is an increased sequencing error due to imprecise base-calling of NGS platforms (Albers et al., 2011). A new technology based on targeted nanopore sequencing with Cas9-guided adapter ligation could be useful for such repetitive regions, as it is able to sequence longer regions, it is not affected by PCR amplification bias and base callers are constantly being improved (Gilpatrick et al., 2020; Xu et al., 2020). However, such technologies are still costly to be used in large scale experiments. Nonetheless, targeted sequencing of repetitive regions has been

successfully applied in plant breeding. Target SSR-seq, a multiplexed amplification of SSRs with a high-throughput sequencing approach, was successfully developed for genetic background selection, gene and QTL mapping, and molecular breeding of cucumber varieties (Yang et al., 2019). Similarly, the AmpSeq-SSR, which could potentially genotype more than a million SSRs at once, was used for the fingerprinting of rice varieties, as well as for mapping resistance genes (Li et al., 2017).

The Amp-seq approach developed in this study has been proven successful for the detection of low-frequency heterozygous mutations in poinsettia, and it can be directly used in the selection of mutants arising from irradiation breeding. Moreover, such an approach could also be applied for different purposes. The current poinsettia breeding programmes focus on improving transportability, shelf life, cutting production yields and temperature tolerance. However, biotic stresses affecting poinsettia have become more common, e.g. *Botrytis*, *Phytophthora*, *Pythium*, *Rhizoctonia*, *Xanthomonas*, *Curtobacterium flaccumfaciens* pv. and *Bemisia tabaci*, and resistance breeding needs to be addressed in poinsettia breeding programmes (van der Ploeg, 2018). Therefore, such targeted high throughput molecular approaches would help improving the current poinsettia breeding programmes, as well as help tackle the future challenges impose by the current climate change scenario.

5.5 Prospect biotechnological applications in poinsettia breeding

Despite its worldwide demand, 75% of cultivated poinsettias in Europe are from red cultivars, while in the German market in 2018, only 20% of the varieties displayed white, pink or bicoloured/scattered coloured bracts (van der Ploeg, 2018). Other uncommon cultivars include curly and rose-shaped bracts, as well as the snow-white and bright pink princettias, but they only reach small niches including younger target groups (van der Ploeg, 2018). Nonetheless, the development of new varieties bearing distinct phenotypes, such as bract colour and shape, will become trivial to meet the demand for different niches. With that in mind, mutation induction in vegetatively propagated ornamentals (e.g. poinsettia) can be of extreme value since it allows for changes in a few characters without changing the remaining genetic background (as compared with conventional breeding) (Broertjes, 1969; Datta and Silva, 2006)

With the availability of genome and transcriptome sequences from several horticultural crops, the application of advanced genetic engineering tools, such as genome editing,

become possible and they can be used to generate variation in different plant traits (Karkute et al., 2017). Genome editing approaches, such as zinc-finger nucleases (ZFNs) (Carroll, 2011), transcription activator-like effector nucleases (TALENs) (Mahfouz et al., 2011) and Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated 9 (CRISPR/Cas9) (Jinek et al., 2012; Cong et al., 2013), have been successfully used to alter different genomic targets in ornamental plants and are recognized as important tools of the improvement of horticultural crops (Xu et al., 2019; Li et al., 2020)

Successful knockout of the *dihydroflavonol-4-reductase-B* (*DFR-B*) gene using the CRISPR/Cas9 system in purple-flowered Japanese morning glory (*Ipomoea nil*) resulted in flowers with white and light purple (due to chimeric mutations) flowers (Watanabe et al., 2017). Another study from the same authors using a white-flowered Japanese morning glory obtained pale yellow flowers by targeting the *carotenoid cleavage dioxygenase 4* (*CCD4*) gene using the CRISPR/Cas9 system (Watanabe et al., 2018). Albino plants were obtained in petunia when using the *Phytoene desaturase* (*PDS*) gene as a target for CRISPR/Cas9 gene editing (Zhang et al., 2016). In transgenic *Chrysanthemum morifolium*, the previously inserted *yellowish-green fluorescent protein* (*YGFP*) gene was used as a target for CRISPR/Cas9 gene editing. This was the first report of gene editing in the species and it is a great achievement since the species is hexaploid, which hinders the applicability of genome editing techniques (Kishi-Kaboshi et al., 2017).

Although genome editing has been widely used for almost a decade, its precision and security, as well its regulation regarding environmental risk assessment (ERA) and management, has been intensively scrutinized (Cho et al., 2014; Agapito-Tenfen et al., 2018; Duensing et al., 2018; Fears and ter Meulen, 2018; Shew et al., 2018). Due to the uncertainties of the technology, the Court of Justice of the European Union (CJEU), in 2018, ruled that genome-edited organisms qualify as products of genetic engineering and hence fall under the scope of the Deliberate Release Directive 2001/18/EC. According to the court ruling, as genome-editing techniques have not yet demonstrated a long safety record in the open field or in a few applications, they cannot be exempted from the rules applying to GMOs. (Callaway, 2018; Confédération paysanne and Others, 2018).

Even though gene editing has an uncertain future about its commercial use, the applicability of such technologies has yet to be tested poinsettia. This current thesis has demonstrated that a mutation in a *GST* gene leads to the appearance of white bracted phenotypes. Similarly, a nonsense mutation in the *flavonoid 3'-hydroxylase* (*F3'H*) gene showed correlations with the appearance of orange coloured bracts (Nitarska et al., 2018). Therefore, creating knock-out mutants in anthocyanin-related genes via CRISPR/Cas9 gene editing could be a promising approach to generate plants with distinct phenotypes and thus extend the portfolio of available cultivars.

5.6 Conclusion

This study aimed to unravel the molecular mechanism responsible for the '*white paradox*' (i.e. the appearance of white bracted phenotypes) in poinsettia, as well as to generate molecular tools and information to improve the breeding programmes for the species. With that in mind, we generated a comprehensive hybrid *de novo* transcriptome from poinsettia bracts, together with the first-hand profiling of gene expression during the process of bract development of red and white varieties. This 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 colour formation in poinsettia, with a focus on the detailed regulatory pathways regulating flavonoid biosynthesis.

Besides the transcriptome data, we identified the *Bract1* gene as an active *GST* gene involved in the expression of anthocyanins in poinsettia bracts, and that a 4 bp deletion in a short repeat within its coding region is most likely the cause of many mutations that lead to a white bract colour (i.e. the '*white paradox*'). This mutation occurs with an unusually high frequency and is presumably an indirect effect of X-ray mutagenesis. Moreover, we were able for the first time to detect low-frequency heterozygous mutation arising from radiation mutagenesis of homozygous varieties, and thus broadening the genetic resources that can be used to generate mutants.

Lastly, we developed an amplicon-seq approach to detect such heterozygous mutations with high accuracy and sensitivity. The approach demonstrated the ability to detect low-frequency mutations at a sensitivity of up to 1 allele in 50 nonmutated alleles, thereby increasing the possibility of high-throughput screenings. This approach may enable an expansion of the genetic resources available for the development of new varieties,

especially those related to colour traits, as well as enabling a reduction in the mutation load in poinsettia gene pools. All in all, this thesis generated powerful and valuable information that can be used in future molecular studies in poinsettia, but that can also be directly applied to breeding programmes.

References

Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., et al. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7, 11706. doi:10.1038/ncomms11706.

Abdelnour-Esquivel, A., Perez, J., Rojas, M., Vargas, W., and Gatica-Arias, A. (2020). Use of gamma radiation to induce mutations in rice (Oryza sativa L.) and the selection of lines with tolerance to salinity and drought. *Vitro Cell. Dev. Biol. - Plant* 56, 88–97. doi:10.1007/s11627-019-10015-5.

Acquaah, G. (2012). *Principles of plant genetics and breeding*. 2nd ed. Hoboken, NJ: Wiley.

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., et al. (2009). Fruit-Surface Flavonoid Accumulation in Tomato Is Controlled by a SlMYB12-Regulated Transcriptional Network. *PLoS Genet.* 5, e1000777. doi:10.1371/journal.pgen.1000777.

Agapito-Tenfen, S. Z., Okoli, A. S., Bernstein, M. J., Wikmark, O.-G., and Myhr, A. I. (2018). Revisiting Risk Governance of GM Plants: The Need to Consider New and Emerging Gene-Editing Techniques. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.01874.

Agati, G., and Tattini, M. (2010). Multiple functional roles of flavonoids in photoprotection. *New Phytol.* 186, 786–793. doi:10.1111/j.1469-8137.2010.03269.x.

Agrawal, G. K., Jwa, N.-S., and Rakwal, R. (2002). A pathogen-induced novel rice (Oryza sativa L.) gene encodes a putative protein homologous to type II glutathione S-transferases. *Plant Sci.* 163, 1153–1160. doi:10.1016/S0168-9452(02)00331-X.

Ahmed, N. U., Park, J.-I., Jung, H.-J., Yang, T.-J., Hur, Y., and Nou, I.-S. (2014). Characterization of dihydroflavonol 4-reductase (DFR) genes and their association with cold and freezing stress in Brassica rapa. *Gene* 550, 46–55. doi:10.1016/j.gene.2014.08.013.

Al-Babili, S., and Bouwmeester, H. J. (2015). Strigolactones, a Novel Carotenoid-Derived Plant Hormone. *Annu. Rev. Plant Biol.* 66, 161–186. doi:10.1146/annurev-arplant-043014-114759.

Albers, C. A., Lunter, G., MacArthur, D. G., McVean, G., Ouwehand, W. H., and Durbin, R. (2011). Dindel: Accurate indel calls from short-read data. *Genome Res.* 21, 961–973. doi:10.1101/gr.112326.110.

Alfenito, M. R., Souer, E., Goodman, C. D., Buell, R., Mol, J., Koes, R., et al. (1998). Functional Complementation of Anthocyanin Sequestration in the Vacuole by Widely Divergent Glutathione S-Transferases. 16.

An, D., Cao, H., Li, C., Humbeck, K., and Wang, W. (2018). Isoform Sequencing and State-of-Art Applications for Unravelling Complexity of Plant Transcriptomes. *Genes* 9, 43. doi:10.3390/genes9010043.

Asen, S. (1958). Anthocyanins in Bracts of Euphorbia Pulcherrima as Revealed by Paper Chromatographic and Spectrophotometric Methods. *Plant Physiol.* 33, 14–7. doi:10.1104/pp.33.1.14.

Au, K. F., Underwood, J. G., Lee, L., and Wong, W. H. (2012). Improving PacBio long read accuracy by short read alignment. *PloS One* 7, e46679. doi:10.1371/journal.pone.0046679.

Babineau, M., Mahmood, K., Mathiassen, S. K., Kudsk, P., and Kristensen, M. (2017). De novo transcriptome assembly analysis of weed Apera spica-venti from seven tissues and growth stages. *BMC Genomics* 18, 128. doi:10.1186/s12864-017-3538-4.

Bado, S., Forster, B. P., Nielen, S., Ali, A. M., Lagoda, P. J. L., Till, B. J., et al. (2015). “Plant Mutation Breeding: Current Progress and Future Assessment,” in *Plant Breeding Reviews*, ed. J. Janick (Hoboken, New Jersey: John Wiley & Sons, Inc.), 23–88. doi:10.1002/9781119107743.ch2.

Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S., and Vivanco, J. M. (2006). The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu. Rev. Plant Biol.* 57, 233–266. doi:10.1146/annurev.arplant.57.032905.105159.

Bassham, D. C. (2015). Pigments on the move. *Nature* 526, 644–645. doi:10.1038/526644a.

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 39, 366–380. doi:10.1111/j.1365-313X.2004.02138.x.

Bempong, M. A., and Sink, K. C. (1968a). MEIOTIC CHROMOSOME ANALYSIS IN THE 42-CHROMOSOME POINSETTIA, EUPHORBIA PULCHERRIMA. *Can. J. Genet. Cytol.* 10, 198–199. doi:10.1139/g68-028.

Bempong, M. A., and Sink, K. C. (1968b). The Karyotype of Poinsettia. *J. Hered.* 59, 259–260. doi:10.1093/oxfordjournals.jhered.a107712.

Ben-Simhon, Z., Judeinstein, S., Trainin, T., Harel-Beja, R., Bar-Ya’akov, I., Borochov-Neori, H., et al. (2015). A “White” Anthocyanin-less Pomegranate (Punica granatum L.) Caused by an Insertion in the Coding Region of the Leucoanthocyanidin Dioxygenase (LDOX; ANS) Gene. *PLOS ONE* 10, e0142777. doi:10.1371/journal.pone.0142777.

Bercu, R. (2016). Comparative Anatomical Aspects of Euphorbia Milli Var. Splendens (bojer Ex Hook.) Ursch & Leandri and Euphorbia Pulcherrima Willd. Ex Klotzsch Species Leaves. *Ann. West Univ. Timiş. Ser. Biol.* 19, 157–166.

Blair, M. W., Fernandez, A. C., Ishitani, M., Moreta, D., Seki, M., Ayling, S., et al. (2011). Construction and EST sequencing of full-length, drought stress cDNA libraries for common beans (Phaseolus vulgaris L.). *BMC Plant Biol.* 11, 171. doi:10.1186/1471-2229-11-171.

Blande, D., Halimaa, P., Tervahauta, A. I., Aarts, M. G. M., and Kärenlampi, S. O. (2017). De novo transcriptome assemblies of four accessions of the metal hyperaccumulator plant Noccaea caerulescens. *Sci. Data* 4, 160131. doi:10.1038/sdata.2016.131.

Bleidorn, C. (2016). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Syst. Biodivers.* 14, 1–8. doi:10.1080/14772000.2015.1099575.

Bleuyard, J.-Y., Gallego, M. E., and White, C. I. (2006). Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair* 5, 1–12. doi:10.1016/j.dnarep.2005.08.017.

Boguski, M., Tolstoshev, C., and Bassett, D. (1994). Gene discovery in dbEST. *Science* 265, 1993–1994. doi:10.1126/science.8091218.

Bolger, M. E., Arsova, B., and Usadel, B. (2017). Plant genome and transcriptome annotations: from misconceptions to simple solutions. *Brief. Bioinform.*, bbw135. doi:10.1093/bib/bbw135.

Braich, S., Baillie, R. C., Jewell, L. S., Spangenberg, G. C., and Cogan, N. O. I. (2019). Generation of a Comprehensive Transcriptome Atlas and Transcriptome Dynamics in Medicinal Cannabis. *Sci. Rep.* 9, 16583. doi:10.1038/s41598-019-53023-6.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18, 630–634. doi:10.1038/76469.

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.

Broertjes, C. (1966). Mutation breeding of chrysanthemums. *Euphytica* 15, 156–162. doi:10.1007/BF00022318.

Broertjes, C. (1969). Mutation breeding of Streptocarpus. *Euphytica* 18, 333–339. doi:10.1007/BF00397780.

Broertjes, C., and Van Harten, A. M. (1988). *Applied mutation breeding for vegetatively propagated crops*. Amsterdam ; New York: Elsevier.

Bruyns, P. V., Mapaya, R. J., and Hedderson, T. J. (2006). A new subgeneric classification for Euphorbia (Euphorbiaceae) in southern Africa based on ITS and psbA-trnH sequence data. *TAXON* 55, 397–420. doi:10.2307/25065587.

Bulgakov, V. P., Avramenko, T. V., and Tsitsiashvili, G. Sh. (2017). Critical analysis of protein signaling networks involved in the regulation of plant secondary metabolism: focus on anthocyanins. *Crit. Rev. Biotechnol.* 37, 685–700. doi:10.3109/07388551.2016.1141391.

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2018). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *bioRxiv*, 420208. doi:10.1101/420208.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.* 55, 611–622. doi:10.1373/clinchem.2008.112797.

Butelli, E., Titta, L., Giorgio, M., Mock, H.-P., Matros, A., Peterek, S., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* 26, 1301–1308. doi:10.1038/nbt.1506.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., et al. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data: DE NOVO NGS-BASED TRANSCRIPTOME ASSEMBLY. *Mol. Ecol. Resour.* 12, 834–845. doi:10.1111/j.1755-0998.2012.03148.x.

Callaway, E. (2018). CRISPR plants now subject to tough GM laws in European Union. *Nature* 560, 16–16. doi:10.1038/d41586-018-05814-6.

Campbell, J., Zhang, H., Giroux, M. J., Feiz, L., Jin, Y., Wang, M., et al. (2012). A mutagenesis-derived broad-spectrum disease resistance locus in wheat. *Theor. Appl. Genet.* 125, 391–404. doi:10.1007/s00122-012-1841-7.

Cao, X., Zhang, F., Yuan, B., Meng, L., Yang, X., and Jiang, J. (2017). De novo transcriptome sequencing and analysis of Euphorbia pekinensis Rupr. and identification of genes involved in diterpenoid biosynthesis. *Plant Gene* 12, 33–42. doi:10.1016/j.plgene.2017.07.001.

Carroll, D. (2011). Genome Engineering With Zinc-Finger Nucleases. *Genetics* 188, 773–782. doi:10.1534/genetics.111.131433.

Ceballos, H., Sánchez, T., Denyer, K., Tofiño, A. P., Rosero, E. A., Dufour, D., et al. (2008). Induction and identification of a small-granule, high-amylose mutant in cassava (Manihot esculenta Crantz). *J. Agric. Food Chem.* 56, 7215–7222. doi:10.1021/jf800603p.

Chakraborty, M., Gangappa, S. N., Maurya, J. P., Sethi, V., Srivastava, A. K., Singh, A., et al. (2019). Functional interrelation of MYC2 and HY5 plays an important role in Arabidopsis seedling development. *Plant J.* 99, 1080–1097. doi:10.1111/tpj.14381.

Chalker-Scott, L. (1999). Environmental Significance of Anthocyanins in Plant Stress Responses. *Photochem. Photobiol.* 70, 1–9. doi:10.1111/j.1751-1097.1999.tb01944.x.

Chandler, V. L., Radicella, J. P., Robbins, T. P., Chen, J., and Turks, D. (1989). Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of B utilizing R genomic sequences. *Plant Cell* 1, 1175–1183. doi:10.1105/tpc.1.12.1175.

Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., et al. (2015). Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol.* 16, 30. doi:10.1186/s13059-015-0596-2.

Chanoca, A., Kovinich, N., Burkel, B., Stecha, S., Bohorquez-Restrepo, A., Ueda, T., et al. (2015). Anthocyanin Vacuolar Inclusions Form by a Microautophagy Mechanism. *Plant Cell* 27, 2545–2559. doi:10.1105/tpc.15.00589.

Chaudhary, J., Deshmukh, R., and Sonah, H. (2019). Mutagenesis Approaches and Their Role in Crop Improvement. *Plants* 8, 467. doi:10.3390/plants8110467.

Cheng, B., Furtado, A., and Henry, R. J. (2017). Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. *GigaScience* 6, 1–13. doi:10.1093/gigascience/gix086.

Cheng, J., Liao, L., Zhou, H., Gu, C., Wang, L., and Han, Y. (2015). A small indel mutation in an anthocyanin transporter causes variegated colouration of peach flowers. *J. Exp. Bot.* 66, 7227–7239. doi:10.1093/jxb/erv419.

Cheynier, 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. PPB* 72, 1–20. doi:10.1016/j.plaphy.2013.05.009.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., et al. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24, 132–141. doi:10.1101/gr.162339.113.

Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C. E., and Burow, M. D. (2014). Comparisons of De Novo Transcriptome Assemblers in Diploid and Polyploid Species Using Peanut (Arachis spp.) RNA-Seq Data. *PLOS ONE* 9, e115055. doi:10.1371/journal.pone.0115055.

Chronopoulou, E., Georgakis, N., Nianiou-Obeidat, I., Madesis, P., Perperopoulou, F., Pouliou, F., et al. (2017). “Plant Glutathione Transferases in Abiotic Stress Response and Herbicide Resistance,” in *Glutathione in Plant Growth, Development, and Stress Tolerance*, eds. M. A. Hossain, M. G. Mostofa, P. Diaz-Vivancos, D. J. Burritt, M. Fujita, and L.-S. P. Tran (Cham: Springer International Publishing), 215–233. doi:10.1007/978-3-319-66682-2_10.

Chung, H.-H., Schwinn, K. E., Ngo, H. M., Lewis, D. H., Massey, B., Calcott, K. E., et al. (2015). Characterisation of betalain biosynthesis in Parakeelya flowers identifies the key biosynthetic gene DOD as belonging to an expanded LigB gene family that is conserved in betalain-producing species. *Front. Plant Sci.* 6. doi:10.3389/fpls.2015.00499.

Clarke, J. L., Klemsdal, S. S., Floistad, E., Hvoslef-Eide, A. K., Haugslien, S., Moe, R., et al. (2006). Genetic Engineering of Poinsettia with the Aim of Enhancing Its Resistance to Poinsettia Mosaic Virus. *Acta Hortic.*, 321–325. doi:10.17660/ActaHortic.2006.722.40.

Clarke, J. L., Spetz, C., Haugslien, S., Xing, S., Dees, M. W., Moe, R., et al. (2008). Agrobacterium tumefaciens-mediated transformation of poinsettia, Euphorbia pulcherrima, with virus-derived hairpin RNA constructs confers resistance to

Poinsettia mosaic virus. *Plant Cell Rep.* 27, 1027–38. doi:10.1007/s00299-008-0526-9.

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana: Floral dip transformation of Arabidopsis. *Plant J.* 16, 735–743. doi:10.1046/j.1365-313x.1998.00343.x.

Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.* 15. doi:10.1371/journal.pbio.2003243.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13. doi:10.1186/s13059-016-0881-8.

Confédération paysanne and Others (2018). Organisms obtained by mutagenesis are GMOs and are, in principle, subject to the obligations laid down by the GMO Directive. Judgment in Case C-528/16. Available at: http://curia.europa.eu/juris/documents.jsf?num=C-528/16.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 819–823. doi:10.1126/science.1231143.

Conn, S., Franco, C., and Zhang, W. (2010). Characterization of anthocyanic vacuolar inclusions in Vitis vinifera L. cell suspension cultures. *Planta* 231, 1343–1360. doi:10.1007/s00425-010-1139-4.

Crick, F. H. (1958). On protein synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.

Cummins, I., Dixon, D. P., Freitag-Pohl, S., Skipsey, M., and Edwards, R. (2011). Multiple roles for plant glutathione transferases in xenobiotic detoxification. *Drug Metab. Rev.* 43, 266–280. doi:10.3109/03602532.2011.552910.

Czemmel, S., Stracke, R., Weisshaar, B., Cordon, N., Harris, N. N., Walker, A. R., et al. (2009). The Grapevine R2R3-MYB Transcription Factor VvMYBF1 Regulates Flavonol Synthesis in Developing Grape Berries. *Plant Physiol.* 151, 1513–1530. doi:10.1104/pp.109.142059.

Datta, S. K. (2018). Breeding of New Ornamental Varieties:Rose. *Curr. Sci.* 114, 1194. doi:10.18520/cs/v114/i06/1194-1206.

Datta, S. K. (2020). Induced mutations: technological advancement for development of new ornamental varieties. *The Nucleus*. doi:10.1007/s13237-020-00310-7.

Datta, S. K., and Silva, J. A. T. da (2006). Role of induced mutagenesis for development of new flower colour and type in ornamentals. *Floric. Ornam. Plant Biotechnol.*, 640–645.

Davies, K. M. (2015). Swapping one red pigment for another. *Nat. Genet.* 47, 5–6. doi:10.1038/ng.3174.

De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., and Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: improved

accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Mol. Ecol. Resour.* 17, 492–507. doi:10.1111/1755-0998.12594.

Debeaujon, I., Peeters, A. J. M., Léon-Kloosterziel, K. M., and Koornneef, M. (2001). The TRANSPARENT TESTA12 Gene of Arabidopsis Encodes a Multidrug Secondary Transporter-like Protein Required for Flavonoid Sequestration in Vacuoles of the Seed Coat Endothelium. *Plant Cell* 13, 853–871. doi:10.1105/tpc.13.4.853.

Debener, T., Lehnackers, H., Arnold, M., and Dangl, J. L. (1991). Identification and molecular mapping of a single Arabidopsis thaliana locus determining resistance to a phytopathogenic Pseudomonas syringae isolate. *Plant J. Cell Mol. Biol.* 1, 289–302. doi:10.1046/j.1365-313X.1991.t01-7-00999.x.

Dixon, D. P., Cummins, I., Cole, D. J., and Edwards, R. (1998). Glutathione-mediated detoxification systems in plants. *Curr. Opin. Plant Biol.* 1, 258–266. doi:10.1016/S1369-5266(98)80114-3.

Dixon, D. P., and Edwards, R. (2010). Glutathione Transferases. *Arab. Book* 8, e0131. doi:10.1199/tab.0131.

Dixon, D. P., Lapthorn, A., and Edwards, R. (2002). Plant glutathione transferases. *Genome Biol.* 3, reviews3004.1. doi:10.1186/gb-2002-3-3-reviews3004.

Dixon, D. P., Skipsey, M., and Edwards, R. (2010). Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry* 71, 338–350. doi:10.1016/j.phytochem.2009.12.012.

Dole, J. M., and Wilkins, H. F. (1991). Vegetative and Reproductive Characteristics of Poinsettia Altered by a Graft-transmissible Agent. *J. Am. Soc. Hortic. Sci.* 116, 307–311. doi:10.21273/JASHS.116.2.307.

Dombrecht, B., Xue, G. P., Sprague, S. J., Kirkegaard, J. A., Ross, J. J., Reid, J. B., et al. (2007). MYC2 Differentially Modulates Diverse Jasmonate-Dependent Functions in *Arabidopsis*. *Plant Cell* 19, 2225–2245. doi:10.1105/tpc.106.048017.

Dong, L., Liu, H., Zhang, J., Yang, S., Kong, G., Chu, J. S. C., et al. (2015). Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genomics* 16, 1039. doi:10.1186/s12864-015-2257-y.

Dong, Z., and Chen, Y. (2013). Transcriptomics: Advances and approaches. *Sci. China Life Sci.* 56, 960–967. doi:10.1007/s11427-013-4557-2.

Draghici, S., Khatri, P., Eklund, A., and Szallasi, Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet.* 22, 101–109. doi:10.1016/j.tig.2005.12.005.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. *Trends Plant Sci.* 15, 573–581. doi:10.1016/j.tplants.2010.06.005.

Duensing, N., Sprink, T., Parrott, W. A., Fedorova, M., Lema, M. A., Wolt, J. D., et al. (2018). Novel Features and Considerations for ERA and Regulation of Crops

Produced by Genome Editing. *Front. Bioeng. Biotechnol.* 6. doi:10.3389/fbioe.2018.00079.

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.

Edwards, R., Dixon, D. P., and Walbot, V. (2000). Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends Plant Sci.* 5, 193–198. doi:10.1016/s1360-1385(00)01601-0.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323, 133–138. doi:10.1126/science.1162986.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* 5, 435–445. doi:10.1038/nrg1348.

Esnault, M.-A., Legue, F., and Chenal, C. (2010). Ionizing radiation: Advances in plant response. *Environ. Exp. Bot.* 68, 231–237. doi:10.1016/j.envexpbot.2010.01.007.

Espley, R. V., Hellens, R. P., Putterill, J., Stevenson, D. E., Kutty-Amma, S., and Allan, A. C. (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* 49, 414–427. doi:10.1111/j.1365-313X.2006.02964.x.

Ewart, L. C., and Walker, D. E. (1960). CHROMOSOME NUMBERS OF POINSETTIA. *J. Hered.* 51, 203–208. doi:10.1093/oxfordjournals.jhered.a106989.

Falcone Ferreyra, M. L., Rius, S. P., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3. doi:10.3389/fpls.2012.00222.

Fan, H., and Chu, J.-Y. (2007). A Brief Review of Short Tandem Repeat Mutation. *Genomics Proteomics Bioinformatics* 5, 7–14. doi:10.1016/S1672-0229(07)60009-6.

FAO, F. and A. O. of the U. N., and IAEA, I. A. E. A. (2019). IAEA Mutant Varieties Database. Available at: https://mvd.iaea.org/ [Accessed April 20, 2020].

Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., et al. (2014). Hyperacidification of Vacuoles by the Combined Action of Two Different P-ATPases in the Tonoplast Determines Flower Color. *Cell Rep.* 6, 32–43. doi:10.1016/j.celrep.2013.12.009.

Fasoli, M., Santo, S. D., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., et al. (2012). The Grapevine Expression Atlas Reveals a Deep Transcriptome Shift Driving the Entire Plant into a Maturation Program. *Plant Cell* 24, 3489–3505. doi:10.1105/tpc.112.100230.

Fears, R., and ter Meulen, V. (2018). Assessing Security Implications of Genome Editing: Emerging Points From an International Workshop. *Front. Bioeng. Biotechnol.* 6. doi:10.3389/fbioe.2018.00034.

Feng, S., Xu, M., Liu, F., Cui, C., and Zhou, B. (2019). Reconstruction of the full-length transcriptome atlas using PacBio Iso-Seq provides insight into the alternative splicing in Gossypium australe. *BMC Plant Biol.* 19, 365. doi:10.1186/s12870-019-1968-7.

Fiedor, J., and Burda, K. (2014). Potential Role of Carotenoids as Antioxidants in Human Health and Disease. *Nutrients* 6, 466–488. doi:10.3390/nu6020466.

Francisco, R. M., Regalado, A., Ageorges, A., Burla, B. J., Bassin, B., Eisenach, C., et al. (2013). ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-Glucosides. *Plant Cell* 25, 1840–1854. doi:10.1105/tpc.112.102152.

Frear, D. S., and Swanson, H. R. (1970). Biosynthesis of S-(4-ethylamino-6-isopropylamino- 2-s-triazino) glutathione: Partial purification and properties of a glutathione S-transferase from corn. *Phytochemistry* 9, 2123–2132. doi:10.1016/S0031-9422(00)85377-7.

Frova, C. (2006). Glutathione transferases in the genomics era: New insights and perspectives. *Biomol. Eng.* 23, 149–169. doi:10.1016/j.bioeng.2006.05.020.

Fungtammasan, A., Ananda, G., Hile, S. E., Su, M. S.-W., Sun, C., Harris, R., et al. (2015). Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res.* 25, 736–749. doi:10.1101/gr.185892.114.

GBIF Secretariat (2019). GBIF Backbone Taxonomy. GBIF Secretariat doi:10.15468/39OMEI.

Gemayel, R., Vinces, M. D., Legendre, M., and Verstrepen, K. J. (2010). Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu. Rev. Genet.* 44, 445–477. doi:10.1146/annurev-genet-072610-155046.

Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., et al. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* 38, 433–438. doi:10.1038/s41587-020-0407-5.

Gisler, B., Salomon, S., and Puchta, H. (2002). The role of double-strand break-induced allelic homologous recombination in somatic plant cells. *Plant J. Cell Mol. Biol.* 32, 277–284. doi:10.1046/j.1365-313x.2002.01421.x.

Golubov, A., Yao, Y., Maheshwari, P., Bilichak, A., Boyko, A., Belzile, F., et al. (2010). Microsatellite Instability in Arabidopsis Increases with Plant Development1[W][OA]. *Plant Physiol.* 154, 1415–1427. doi:10.1104/pp.110.162933.

Gomez, C., Conejero, G., Torregrosa, L., Cheynier, V., Terrier, N., and Ageorges, A. (2011). In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST: Anthocyanin trafficking in grapevine. *Plant J.* 67, 960–970. doi:10.1111/j.1365-313X.2011.04648.x.

Gomez, C., Terrier, N., Torregrosa, L., Vialet, S., Fournier-Level, A., Verriès, C., et al. (2009). Grapevine MATE-Type Proteins Act as Vacuolar H^+-Dependent Acylated

Anthocyanin Transporters. *Plant Physiol.* 150, 402–415. doi:10.1104/pp.109.135624.

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J.* 53, 814–827. doi:10.1111/j.1365-313X.2007.03373.x.

Goodman, C. D., Casati, P., and Walbot, V. (2004). A Multidrug Resistance–Associated Protein Involved in Anthocyanin Transport in *Zea mays*. *Plant Cell* 16, 1812–1826. doi:10.1105/tpc.022574.

Govaerts, R., Frodin, D. G., and Radcliffe-Smith, A. (2000). World checklist and bibliography of Euphorbiaceae (with Pandaceae). 4,1.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi:10.1038/nbt.1883.

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., et al. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics* 164, 731–740.

Grotewold, E. (2006a). The Genetics and Biochemistry of Floral Pigments. *Annu. Rev. Plant Biol.* 57, 761–780. doi:10.1146/annurev.arplant.57.032905.105248.

Grotewold, E. ed. (2006b). *The science of flavonoids*. New York: Springer.

Grotewold, E., Chamberlin, M., Snook, M., Siame, B., Butler, L., Swenson, J., et al. (1998). Engineering secondary metabolism in maize cells by ectopic expression of transcription factors. *Plant Cell* 10, 721–740.

Grotewold, E., and Davies, K. (2008). Trafficking and Sequestration of Anthocyanins. *Nat. Prod. Commun.* 3, 1934578X0800300. doi:10.1177/1934578X0800300806.

Gu, Z., Chen, H., Yang, R., and Ran, M. (2018). Identification of DFR as a promoter of anthocyanin accumulation in poinsettia (Euphorbia pulcherrima , willd. ex Klotzsch) bracts under short-day conditions. *Sci. Hortic.* 236, 158–165. doi:10.1016/j.scienta.2018.03.032.

Gullner, G., Komives, T., Király, L., and Schröder, P. (2018). Glutathione S-Transferase Enzymes in Plant-Pathogen Interactions. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.01836.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi:10.1038/nprot.2013.084.

Haas, B. J., and Zody, M. C. (2010). Advancing RNA-Seq analysis. *Nat. Biotechnol.* 28, 421–423. doi:10.1038/nbt0510-421.

Hall, T. A., Hall, T., Hall, T. A., Hall, T., Hall, T. A., Hall, T. A., et al. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. doi:10.14601/Phytopathol_Mediterr-14998u1.29.

Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., et al. (2008). Single-molecule DNA sequencing of a viral genome. *Science* 320, 106–109. doi:10.1126/science.1150427.

Harvaux, M., and Kloppstech, K. (2001). The protective functions of carotenoid and flavonoid pigments against excess visible radiation at chilling temperature investigated in Arabidopsis npq and tt mutants. *Planta* 213, 953–966. doi:10.1007/s004250100572.

Hashimoto, H., Uragami, C., and Cogdell, R. J. (2016). “Carotenoids and Photosynthesis,” in *Carotenoids in Nature: Biosynthesis, Regulation and Function* Subcellular Biochemistry., ed. C. Stange (Cham: Springer International Publishing), 111–139. doi:10.1007/978-3-319-39126-7_4.

Hatlestad, G. J., Akhavan, N. A., Sunnadeniya, 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., Sunnadeniya, 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.

He, G., Guan, C.-N., Chen, Q.-X., Gou, X.-J., Liu, W., Zeng, Q.-Y., et al. (2016). Genome-Wide Analysis of the Glutathione S-Transferase Gene Family in Capsella rubella: Identification, Expression, and Biochemical Functions. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01325.

Heisler, M. G., Atkinson, A., Bylstra, Y. H., Walsh, R., and Smyth, D. R. (2001). SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. *Development* 128, 1089–1098.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* 62, 2465–2483. doi:10.1093/jxb/erq442.

Hoang, N. V., Furtado, A., Mason, P. J., Marquardt, A., Kasirajan, L., Thirugnanasambandam, P. P., et al. (2017). A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* 18, 395. doi:10.1186/s12864-017-3757-8.

Hölzer, M., and Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. *GigaScience* 8. doi:10.1093/gigascience/giz039.

Horn, J. W., van Ee, B. W., Morawetz, J. J., Riina, R., Steinmann, V. W., Berry, P. E., et al. (2012). Phylogenetics and the evolution of major structural characters in the

giant genus Euphorbia L. (Euphorbiaceae). *Mol. Phylogenet. Evol.* 63, 305–26. doi:10.1016/j.ympev.2011.12.022.

Hrdlickova, R., Toloue, M., and Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *WIREs RNA* 8, e1364. doi:10.1002/wrna.1364.

Hsieh, K., and Huang, A. H. C. (2007). Tapetosomes in Brassica Tapetum Accumulate Endoplasmic Reticulum–Derived Flavonoids and Alkanes for Delivery to the Pollen Surface. *Plant Cell* 19, 582–596. doi:10.1105/tpc.106.049049.

Hsieh, P.-H., Oyang, Y.-J., and Chen, C.-Y. (2019). Effect of de novo transcriptome assembly on transcript quantification. *Sci. Rep.* 9, 1–12. doi:10.1038/s41598-019-44499-3.

Hu, B., Zhao, J., Lai, B., Qin, Y., Wang, H., and Hu, G. (2016). LcGST4 is an anthocyanin-related glutathione S-transferase gene in Litchi chinensis Sonn. *Plant Cell Rep.* 35, 831–843. doi:10.1007/s00299-015-1924-4.

Ibrahim, R., Ahmad, Z., Salleh, S., Hassan, A. A., and Ariffin, S. (2018). “Mutation Breeding in Ornamentals,” in *Ornamental Crops* Handbook of Plant Breeding., ed. J. Van Huylenbroeck (Cham: Springer International Publishing), 175–211. doi:10.1007/978-3-319-90698-0_8.

Islam, M. A., Lütken, H., Haugslien, S., Blystad, D.-R., Torre, S., Rolcik, J., et al. (2013). Overexpression of the AtSHI gene in poinsettia, Euphorbia pulcherrima, results in compact plants. *PloS One* 8, 53377. doi:10.1371/journal.pone.0053377.

Islam, S., Sajib, S. D., Jui, Z. S., Arabia, S., Islam, T., and Ghosh, A. (2019). Genome-wide identification of glutathione S-transferase gene family in pepper, its classification, and expression profiling under different anatomical and environmental conditions. *Sci. Rep.* 9, 9101. doi:10.1038/s41598-019-45320-x.

Jaakola, L., and Hohtola, A. (2010). Effect of latitude on flavonoid biosynthesis in plants. *Plant Cell Environ.* 33, 1239–1247. doi:10.1111/j.1365-3040.2010.02154.x.

Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., and Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *Nat. Methods* 12, 351–356. doi:10.1038/nmeth.3290.

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi:10.1038/nbt.4060.

Jain, M., Tyson, J. R., Loose, M., Ip, C. L. C., Eccles, D. A., O’Grady, J., et al. (2017). MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. *F1000Research* 6, 760. doi:10.12688/f1000research.11354.1.

Jain, P., Krishnan, N. M., and Panda, B. (2013). Augmenting transcriptome assembly by combining *de novo* and genome-guided tools. *PeerJ* 1, e133. doi:10.7717/peerj.133.

Jantzen, F., Wozniak, N., Kappel, C., Sicard, A., and Lenhard, M. (2019). A high-throughput amplicon-based method for estimating outcrossing rates. *Plant Methods* 15, 47. doi:10.1186/s13007-019-0433-9.

Jiang, S., Chen, M., He, N., Chen, X., Wang, N., Sun, Q., et al. (2019). MdGSTF6, activated by MdMYB1, plays an essential role in anthocyanin accumulation in apple. *Hortic. Res.* 6, 40. doi:10.1038/s41438-019-0118-6.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821. doi:10.1126/science.1225829.

Jo, Y. D., and Kim, J.-B. (2019). Frequency and Spectrum of Radiation-Induced Mutations Revealed by Whole-Genome Sequencing Analyses of Plants. *Quantum Beam Sci.* 3, 7. doi:10.3390/qubs3020007.

Jordan, B. R. (2018). “Plant Pigments and Protection against UV-B Radiation,” in *Annual Plant Reviews online* (American Cancer Society), 275–292. doi:10.1002/9781119312994.apr0139.

Kagale, S., Nixon, J., Khedikar, Y., Pasha, A., Provart, N. J., Clarke, W. E., et al. (2016). The developmental transcriptome atlas of the biofuel crop Camelina sativa. *Plant J. Cell Mol. Biol.* 88, 879–894. doi:10.1111/tpj.13302.

Kannan, S., Hui, J., Mazooji, K., Pachter, L., and Tse, D. (2016). Shannon: An Information-Optimal de Novo RNA-Seq Assembler. *bioRxiv*, 039230. doi:10.1101/039230.

Kannangara, C. G., and Hansson, M. (1998). Arrest of chlorophyll accumulation prior to anthocyanin formation in Euphorbia pulcherrima. *Plant Physiol. Biochem.* 36, 843–848. doi:10.1016/S0981-9428(99)80001-1.

Karakach, T. K., Flight, R. M., Douglas, S. E., and Wentzell, P. D. (2010). An introduction to DNA microarrays for gene expression analysis. *Chemom. Intell. Lab. Syst.* 104, 28–52. doi:10.1016/j.chemolab.2010.04.003.

Karkute, S. G., Singh, A. K., Gupta, O. P., Singh, P. M., and Singh, B. (2017). CRISPR/Cas9 Mediated Genome Engineering for Improvement of Horticultural Crops. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.01635.

Kazama, Y., Hirano, T., Saito, H., Liu, Y., Ohbu, S., Hayashi, Y., et al. (2011). Characterization of highly efficient heavy-ion mutagenesis in Arabidopsis thaliana. *BMC Plant Biol.* 11, 161. doi:10.1186/1471-2229-11-161.

Kazan, K., and Manners, J. M. (2013). MYC2: The Master in Action. *Mol. Plant* 6, 686–703. doi:10.1093/mp/sss128.

Kevan, P. G., and Baker, H. G. (1983). Insects as Flower Visitors and Pollinators. *Annu. Rev. Entomol.* 28, 407–453. doi:10.1146/annurev.en.28.010183.002203.

Khalil, F., Naiyan, X., Tayyab, M., and Pinghua, C. (2018). Screening of EMS-Induced Drought-Tolerant Sugarcane Mutants Employing Physiological, Molecular and Enzymatic Approaches. *Agronomy* 8, 226. doi:10.3390/agronomy8100226.

Kishi-Kaboshi, M., Aida, R., and Sasaki, K. (2017). Generation of Gene-Edited Chrysanthemum morifolium Using Multicopy Transgenes as Targets and Markers. *Plant Cell Physiol.* 58, 216–226. doi:10.1093/pcp/pcw222.

Kitamura, S., Akita, Y., Ishizaka, H., Narumi, I., and Tanaka, A. (2012). Molecular characterization of an anthocyanin-related glutathione S-transferase gene in cyclamen. *J. Plant Physiol.* 169, 636–642. doi:10.1016/j.jplph.2011.12.011.

Kitamura, S., Shikazono, N., and Tanaka, A. (2004). *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *Plant J.* 37, 104–114. doi:10.1046/j.1365-313X.2003.01943.x.

Knoll, A., Fauser, F., and Puchta, H. (2014). DNA recombination in somatic plant cells: mechanisms and evolutionary consequences. *Chromosome Res.* 22, 191–201. doi:10.1007/s10577-014-9415-y.

Kobayashi, R. (2012). Free branching poinsettia. Available at: https://patents.google.com/patent/WO2012177688A1/en [Accessed March 17, 2020].

Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci.* 10, 236–242. doi:10.1016/j.tplants.2005.03.002.

Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., et al. (2012). Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700. doi:10.1038/nbt.2280.

Koseki, M., Goto, K., Masuta, C., and Kanazawa, A. (2005). The Star-type Color Pattern in Petunia hybrida 'Red Star' Flowers is Induced by Sequence-Specific Degradation of Chalcone Synthase RNA. *Plant Cell Physiol.* 46, 1879–1883. doi:10.1093/pcp/pci192.

Kou, M., Liu, Y., Li, Z., Zhang, Y., Tang, W., Yan, H., et al. (2019). A novel glutathione S-transferase gene from sweetpotato, IbGSTF4, is involved in anthocyanin sequestration. *Plant Physiol. Biochem.* 135, 395–403. doi:10.1016/j.plaphy.2018.12.028.

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 278. doi:10.1186/s13059-019-1910-1.

Kuang, X., Sun, S., Wei, J., Li, Y., and Sun, C. (2019). Iso-Seq analysis of the Taxus cuspidata transcriptome reveals the complexity of Taxol biosynthesis. *BMC Plant Biol.* 19, 210. doi:10.1186/s12870-019-1809-8.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi:10.1093/molbev/msy096.

Kumar, S., and Trivedi, P. K. (2018). Glutathione S-Transferases: Role in Combating Abiotic Stresses Including Arsenic Detoxification in Plants. *Front. Plant Sci.* 9, 751. doi:10.3389/fpls.2018.00751.

Kunter, B., Bas, M., Kantoglu, Y., and Burak, M. (2012). “Mutation breeding of sweet cherry (Prunus avium L.) var. 0900 Ziraat.,” in *Plant mutation breeding and biotechnology*, eds. Q. Y. Shu, B. P. Forster, and H. Nakagawa (Wallingford: CABI), 453–459. doi:10.1079/9781780640853.0453.

Lack, H. W. (2011). The discovery, naming and typification of Euphorbia pulcherrima (Euphorbiaceae). *Willdenowia* 41, 301–309. doi:10.3372/wi.41.41212.

Lallement, P.-A., Brouwer, B., Keech, O., Hecker, A., and Rouhier, N. (2014). The still mysterious roles of cysteine-containing glutathione transferases in plants. *Front. Pharmacol.* 5. doi:10.3389/fphar.2014.00192.

Landi, M., Tattini, M., and Gould, K. S. (2015). Multiple functional roles of anthocyanins in plant-environment interactions. *Environ. Exp. Bot.* 119, 4–17. doi:10.1016/j.envexpbot.2015.05.012.

Larsen, E. S., Alfenito, M. R., Briggs, W. R., and Walbot, V. (2003). A carnation anthocyanin mutant is complemented by the glutathione S-transferases encoded by maize Bz2 and petunia An9. *Plant Cell Rep.* 21, 900–904. doi:10.1007/s00299-002-0545-x.

Lee, I.-M. (2000). Phytoplasma Casts a Magic Spell that Turns the Fair Poinsettia into a Christmas Showpiece. *Plant Health Prog.* 1, 16. doi:10.1094/PHP-2000-0914-01-RV.

Lee, I.-M., Klopmeyer, M., Bartoszyk, I. M., Gundersen-Rindal, D. E., Chou, T.-S., Thomson, K. L., et al. (1997). Phytoplasma induced free-branching in commercial poinsettia cultivars. *Nat. Biotechnol.* 15, 178–182. doi:10.1038/nbt0297-178.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242-245. doi:10.1093/nar/gkw290.

Li, F.-D., Tong, W., Xia, E.-H., and Wei, C.-L. (2019a). Optimized sequencing depth and de novo assembler for deeply reconstructing the transcriptome of the tea plant, an economically important plant species. *BMC Bioinformatics* 20, 553. doi:10.1186/s12859-019-3166-x.

Li, L., Fang, Z., Zhou, J., Chen, H., Hu, Z., Gao, L., et al. (2017). An accurate and efficient method for large-scale SSR genotyping and applications. *Nucleic Acids Res.* 45, e88. doi:10.1093/nar/gkx093.

Li, Q., Sapkota, M., and van der Knaap, E. (2020). Perspectives of CRISPR/Cas-mediated cis -engineering in horticulture: unlocking the neglected potential for crop improvement. *Hortic. Res.* 7, 1–11. doi:10.1038/s41438-020-0258-8.

Li, T., Fang, Z., Peng, H., Zhou, J., Liu, P., Wang, Y., et al. (2019b). Application of high-throughput amplicon sequencing-based SSR genotyping in genetic background screening. *BMC Genomics* 20, 444. doi:10.1186/s12864-019-5800-4.

Li, X., Song, Y., Century, K., Straight, S., Ronald, P., Dong, X., et al. (2001). A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant J.* 27, 235–242. doi:10.1046/j.1365-313x.2001.01084.x.

Liang, P., and Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967–971. doi:10.1126/science.1354393.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., et al. (2010). An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. *Plant J.* 63, 86–99. doi:https://doi.org/10.1111/j.1365-313X.2010.04222.x.

Licciardello, C., D'Agostino, N., Traini, A., Recupero, G., Frusciante, L., and Chiusano, M. (2014). Characterization of the glutathione S-transferase gene family through ESTs and expression analyses within common and pigmented cultivars of Citrus sinensis (L.) Osbeck. *BMC Plant Biol.* 14, 39. doi:10.1186/1471-2229-14-39.

Lima, L., Sinaimeri, B., Sacomoto, G., Lopez-Maestre, H., Marchet, C., Miele, V., et al. (2017). Playing hide and seek with repeats in local and global de novo transcriptome assembly of short RNA-seq reads. *Algorithms Mol. Biol. AMB* 12. doi:10.1186/s13015-017-0091-2.

Ling, J.-T., Sauve, R., and Gawel, N. (1997). Identification of Poinsettia Cultivars Using RAPD Markers. *HortScience* 32, 122–124. doi:10.21273/HORTSCI.32.1.122.

Liu, J., Li, G., Chang, Z., Yu, T., Liu, B., McMullen, R., et al. (2016). BinPacker: Packing-Based De Novo Transcriptome Assembly from RNA-seq Data. *PLOS Comput. Biol.* 12, e1004772. doi:10.1371/journal.pcbi.1004772.

Liu, Y., Hou, H., Jiang, X., Wang, P., Dai, X., Chen, W., et al. (2018a). A WD40 Repeat Protein from Camellia sinensis Regulates Anthocyanin and Proanthocyanidin Accumulation through the Formation of MYB–bHLH–WD40 Ternary Complexes. *Int. J. Mol. Sci.* 19, 1686. doi:10.3390/ijms19061686.

Liu, Y., Qi, Y., Zhang, A., Wu, H., Liu, Z., and Ren, X. (2019). Molecular cloning and functional characterization of AcGST1, an anthocyanin-related glutathione S-transferase gene in kiwifruit (Actinidia chinensis). *Plant Mol. Biol.* 100, 451–465. doi:10.1007/s11103-019-00870-6.

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018b). Anthocyanin Biosynthesis and Degradation Mechanisms in Solanaceous Vegetables: A Review. *Front. Chem.* 6, 52. doi:10.3389/fchem.2018.00052.

Lockhart, D. J., and Winzeler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature* 405, 827–836. doi:10.1038/35015701.

Luo, H., Dai, C., Li, Y., Feng, J., Liu, Z., and Kang, C. (2018). Reduced Anthocyanins in Petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. *J. Exp. Bot.* 69, 2595–2608. doi:10.1093/jxb/ery096.

Luo, P., Ning, G., Wang, Z., Shen, Y., Jin, H., Li, P., et al. (2016). Disequilibrium of Flavonol Synthase and Dihydroflavonol-4-Reductase Expression Associated Tightly to White vs. Red Color Flower Formation in Plants. *Front. Plant Sci.* 6. doi:10.3389/fpls.2015.01257.

Mahfouz, M. M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., and Zhu, J.-K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci.* 108, 2623–2628. doi:10.1073/pnas.1019533108.

Malone, J. H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* 9, 34. doi:10.1186/1741-7007-9-34.

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J.-M., et al. (2007). The *Arabidopsis* MATE Transporter TT12 Acts as a Vacuolar Flavonoid/H^+-Antiporter Active in Proanthocyanidin-Accumulating Cells of the Seed Coat. *Plant Cell* 19, 2023–2038. doi:10.1105/tpc.106.046029.

Marrs, K. A., Alfenito, M. R., Lloyd, A. M., and Walbot, V. (1995). A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature* 375, 397–400. doi:10.1038/375397a0.

Martin, J. A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12, 671–682. doi:10.1038/nrg3068.

Mba, C. (2013). Induced Mutations Unleash the Potentials of Plant Genetic Resources for Food and Agriculture. *Agronomy* 3, 200–231. doi:10.3390/agronomy3010200.

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The Arabidopsis Transcription Factor MYB12 Is a Flavonol-Specific Regulator of Phenylpropanoid Biosynthesis. *Plant Physiol.* 138, 1083–1096. doi:10.1104/pp.104.058032.

Mol, J., Grotewold, E., and Koes, R. (1998). How genes paint flowers and seeds. *Trends Plant Sci.* 3, 212–217. doi:10.1016/S1360-1385(98)01242-4.

Mondal, S., Badigannavar, A. M., and D'Souza, S. F. (2011). Induced variability for fatty acid profile and molecular characterization of high oleate mutant in cultivated groundnut (Arachis hypogaea L.). *Plant Breed.* 130, 242–247. doi:10.1111/j.1439-0523.2010.01787.x.

Moons, A. (2005). "Regulatory and Functional Interactions of Plant Growth Regulators and Plant Glutathione S-Transferases (GSTs)," in *Vitamins & Hormones* Plant Hormones., ed. G. Litwack (Academic Press), 155–202. doi:10.1016/S0083-6729(05)72005-7.

Morita, R., Kusaba, M., Iida, S., Yamaguchi, H., Nishio, T., and Nishimura, M. (2009). Molecular characterization of mutations induced by gamma irradiation in rice. *Genes Genet. Syst.* 84, 361–370. doi:10.1266/ggs.84.361.

Morita, Y., Saito, R., Ban, Y., Tanikawa, N., Kuchitsu, K., Ando, T., et al. (2012). Tandemly arranged chalcone synthase A genes contribute to the spatially regulated expression of siRNA and the natural bicolor floral phenotype in Petunia hybrida: Endogenous siRNA and bicolored flowers of Petunia. *Plant J.* 70, 739–749. doi:10.1111/j.1365-313X.2012.04908.x.

Moustaka, J., Panteris, E., Adamakis, I.-D. S., Tanou, G., Giannakoula, A., Eleftheriou, E. P., et al. (2018). High anthocyanin accumulation in poinsettia leaves is accompanied by thylakoid membrane unstacking, acting as a photoprotective

mechanism, to prevent ROS formation. *Environ. Exp. Bot.* 154, 44–55. doi:10.1016/j.envexpbot.2018.01.006.

Moyer, L. S. (1934). Electrophoresis of Latex and Chromosome Numbers of Poinsettias. *Bot. Gaz.* 95, 678–685.

Mueller, L. A., Goodman, C. D., Silady, R. A., and Walbot, V. (2000). AN9, a Petunia Glutathione *S* -Transferase Required for Anthocyanin Sequestration, Is a Flavonoid-Binding Protein. *Plant Physiol.* 123, 1561–1570. doi:10.1104/pp.123.4.1561.

Nagano, S., Hirao, T., Takashima, Y., Matsushita, M., Mishima, K., Takahashi, M., et al. (2020). SNP Genotyping with Target Amplicon Sequencing Using a Multiplexed Primer Panel and Its Application to Genomic Prediction in Japanese Cedar, Cryptomeria japonica (L.f.) D.Don. *Forests* 11, 898. doi:10.3390/f11090898.

Nakasugi, K., Crowhurst, R., Bally, J., and Waterhouse, P. (2014). Combining Transcriptome Assemblies from Multiple De Novo Assemblers in the Allo-Tetraploid Plant Nicotiana benthamiana. *PLOS ONE* 9, e91776. doi:10.1371/journal.pone.0091776.

Nambara, E., and Marion-Poll, A. (2005). Abscisic acid biosynthesis and catabolism. *Annu. Rev. Plant Biol.* 56, 165–185. doi:10.1146/annurev.arplant.56.032604.144046.

Nielsen, M. D., Farestveit, B., and Andersen, A. S. (2003). Adventitious Shoot Development from Decapitated Plants of Peri- clinal Chimeric Poinsettia Plants (Euphorbia pulcherrima Willd ex Klotsch). *Europ.J.Hort.Sci.* 68, 161–168.

Nisar, N., Li, L., Lu, S., Khin, N. C., and Pogson, B. J. (2015). Carotenoid Metabolism in Plants. *Mol. Plant* 8, 68–82. doi:10.1016/j.molp.2014.12.007.

Nishihara, M., Yamada, E., Saito, M., Fujita, K., Takahashi, H., and Nakatsuka, T. (2014). Molecular characterization of mutations in white-flowered torenia plants. *BMC Plant Biol.* 14, 86. doi:10.1186/1471-2229-14-86.

Nitarska, D., Stefanini, C., Haselmair-Gosch, C., Miosic, S., Walliser, B., Mikulic-Petkovsek, M., et al. (2018). 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.* 18, 216. doi:10.1186/s12870-018-1424-0.

Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11, 378. doi:10.1186/1471-2105-11-378.

Nozue, M., and Yasuda, H. (1985). Occurrence of anthocyanoplasts in cell suspension cultures of sweet potato. *Plant Cell Rep.* 4, 252–255. doi:10.1007/BF00269370.

Oakley, A. (2011). Glutathione transferases: a structural perspective. *Drug Metab. Rev.* 43, 138–151. doi:10.3109/03602532.2011.558093.

Okamura, M., Hase, Y., Furusawa, Y., and Tanaka, A. (2015). Tissue-dependent somaclonal mutation frequencies and spectra enhanced by ion beam irradiation in chrysanthemum. *Euphytica* 202, 333–343. doi:10.1007/s10681-014-1220-3.

Okamura, M., Nakayama, M., Umemoto, N., Cano, E. A., Hase, Y., Nishizaki, Y., et al. (2013). Crossbreeding of a metallic color carnation and diversification of the peculiar coloration by ion-beam irradiation. *Euphytica* 191, 45–56. doi:10.1007/s10681-012-0859-x.

Okamura, M., Yasuno, N., Ohtsuka, M., Tanaka, A., Shikazono, N., and Hase, Y. (2003). Wide variety of flower-color and -shape mutants regenerated from leaf cultures irradiated with ion beams. *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.* 206, 574–578. doi:10.1016/S0168-583X(03)00835-8.

Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., et al. (2016). Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnol. Biotechnol. Equip.* 30, 1–16. doi:10.1080/13102818.2015.1087333.

Orel, N., Kyryk, A., and Puchta, H. (2003). Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. *Plant J.* 35, 604–612. doi:10.1046/j.1365-313X.2003.01832.x.

Ortigosa, A., Fonseca, S., Franco-Zorrilla, J. M., Fernández-Calvo, P., Zander, M., Lewsey, M. G., et al. (2020). The JA-pathway MYC transcription factors regulate photomorphogenic responses by targeting HY5 gene expression. *Plant J.*, tpj.14618. doi:10.1111/tpj.14618.

Ozsolak, F., and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98. doi:10.1038/nrg2934.

Palma-Silva, C., Ferro, M., Bacci, M., and Turchetto-Zolet, A. C. (2016). De novo assembly and characterization of leaf and floral transcriptomes of the hybridizing bromeliad species (Pitcairnia spp.) adapted to Neotropical Inselbergs. *Mol. Ecol. Resour.* 16, 1012–1022. doi:10.1111/1755-0998.12504.

Palovaara, J., Saiga, S., Wendrich, J. R., van ‘t Wout Hofland, N., van Schayck, J. P., Hater, F., et al. (2017). Transcriptome dynamics revealed by a gene expression atlas of the early Arabidopsis embryo. *Nat. Plants* 3, 894–904. doi:10.1038/s41477-017-0035-3.

Pâques, F., Leung, W.-Y., and Haber, J. E. (1998). Expansions and Contractions in a Tandem Repeat Induced by Double-Strand Break Repair. *Mol. Cell. Biol.* 18, 2045–2054.

Park, K.-R., and Jansen, R. K. (2007). A Phylogeny of Euphorbieae Subtribe Euphorbiinae (Euphorbiaceae) Based on Molecular Data. *J. Plant Biol.* 50, 644–649. doi:10.1007/BF03030608.

Parks, E. J., and Moyer, J. W. (2004). Evaluation of AFLP in poinsettia: Polymorphism selection, analysis, and cultivar identification. v. 129.

Pathirana, R. (1992). Gamma Ray-Induced Field Tolerance to Phytophthora Blight in Sesame. *Plant Breed.* 108, 314–319. doi:10.1111/j.1439-0523.1992.tb00136.x.

Pearson, C. E., Edamura, K. N., and Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6, 729–742. doi:10.1038/nrg1689.

Peer, W. A., and Murphy, A. S. (2007). Flavonoids and auxin transport: modulators or regulators? *Trends Plant Sci.* 12, 556–563. doi:10.1016/j.tplants.2007.10.003.

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2010). "IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler," in *Research in Computational Molecular Biology* Lecture Notes in Computer Science., ed. B. Berger (Berlin, Heidelberg: Springer Berlin Heidelberg), 426–440. doi:10.1007/978-3-642-12683-3_28.

Pérez-Díaz, R., Madrid-Espinoza, J., Salinas-Cornejo, J., González-Villanueva, E., and Ruiz-Lara, S. (2016). Differential Roles for VviGST1, VviGST3, and VviGST4 in Proanthocyanidin and Anthocyanin Transport in Vitis vinífera. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01166.

Pérez-Nicolás, M. L., Colinas-León, M. T., Gómez-Alanis, M. C., Flores-Espinosa, C., Alia-Tejacal, I., and Peña-Ortega, M. G. (2020). Geographic distribution of species of Euphorbia sect. Poinsettia with potential ornamental use in Mexico. *Polibotánica*, 165–189. doi:10.18387/polibotanica.50.12.

Peters, D. J., and Constabel, C. P. (2002). Molecular analysis of herbivore-induced condensed tannin synthesis: cloning and expression of dihydroflavonol reductase from trembling aspen (Populus tremuloides). *Plant J.* 32, 701–712. doi:10.1046/j.1365-313X.2002.01458.x.

Petroni, K., and Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* 181, 219–229. doi:10.1016/j.plantsci.2011.05.009.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, 45e–445. doi:10.1093/nar/29.9.e45.

Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30, 10.

Pitzschke, A., and Persak, H. (2012). Poinsettia protoplasts - a simple, robust and efficient system for transient gene expression studies. *Plant Methods* 8, 14. doi:10.1186/1746-4811-8-14.

Pomar, F., and Ros Barceló, A. (2007). Are red leaves photosynthetically active? *Biol. Plant.* 51, 799–800. doi:10.1007/s10535-007-0164-z.

Pourcel, L., Irani, N. G., Lu, Y., Riedl, K., Schwartz, S., and Grotewold, E. (2010). The Formation of Anthocyanic Vacuolar Inclusions in Arabidopsis thaliana and Implications for the Sequestration of Anthocyanin Pigments. *Mol. Plant* 3, 78–90. doi:10.1093/mp/ssp071.

Poustka, F., Irani, N. G., Feller, A., Lu, Y., Pourcel, L., Frame, K., et al. (2007). A Trafficking Pathway for Anthocyanins Overlaps with the Endoplasmic Reticulum-to-Vacuole Protein-Sorting Route in Arabidopsis and Contributes to the Formation of Vacuolar Inclusions. *Plant Physiol.* 145, 1323–1335. doi:10.1104/pp.107.105064.

Prenner, G., Cacho, N. I., Baum, D., and Rudall, P. J. (2010). Is LEAFY a useful marker gene for the flower–inflorescence boundary in the Euphorbia cyathium? *J. Exp. Bot.* 62, 345–50. doi:10.1093/jxb/erq275.

Prenner, G., Hopper, S. D., and Rudall, P. J. (2008). Pseudanthium development in Calycopeplus paucifolius, with particular reference to the evolution of the cyathium in Euphorbieae (Euphorbiaceae–Malpighiales). *Aust. Syst. Bot.* 21, 153–61.

Puchta, H. (1999). Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. *Genetics* 152, 1173–1181.

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14. doi:10.1093/jxb/eri025.

Pushkarev, D., Neff, N. F., and Quake, S. R. (2009). Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 27, 847–850. doi:10.1038/nbt.1561.

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics* 13, 341. doi:10.1186/1471-2164-13-341.

Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* 13, 278–289. doi:10.1016/j.gpb.2015.08.002.

Richard, G.-F., and Pâques, F. (2000). Mini- and microsatellite expansions: the recombination connection. *EMBO Rep.* 1, 122–126. doi:10.1093/embo-reports/kvd031.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., et al. (2010). De novo assembly and analysis of RNA-seq data. *Nat. Methods* 7, 909–912. doi:10.1038/nmeth.1517.

Rocha Latado, R., Hernán Adames, A., and Tulmann Neto, A. (2004). In vitro Mutation of Chrysanthemum (Dendranthema grandiflora Tzvelev) with Ethylmethanesulphonate (EMS) in Immature Floral Pedicels. *Plant Cell Tissue Organ Cult.* 77, 103–106. doi:10.1023/B:TICU.0000016481.18358.55.

Roldán-Arjona, T., and Ariza, R. R. (2009). Repair and tolerance of oxidative DNA damage in plants. *Mutat. Res. Mutat. Res.* 681, 169–179. doi:10.1016/j.mrrev.2008.07.003.

Roxas, V. P., Lodhi, S. A., Garrett, D. K., Mahan, J. R., and Allen, R. D. (2000). Stress tolerance in transgenic tobacco seedlings that overexpress glutathione S-transferase/glutathione peroxidase. *Plant Cell Physiol.* 41, 1229–1234. doi:10.1093/pcp/pcd051.

Roychowdhury, R., and Tah, J. (2013). “Mutagenesis—A Potential Approach for Crop Improvement,” in *Crop Improvement: New Approaches and Modern Techniques*, eds. K. R. Hakeem, P. Ahmad, and M. Ozturk (Boston, MA: Springer US), 149–187. doi:10.1007/978-1-4614-7028-1_4.

Ruiz-Sola, M. Á., and Rodríguez-Concepción, M. (2012). Carotenoid Biosynthesis in Arabidopsis: A Colorful Pathway. *Arab. Book Am. Soc. Plant Biol.* 10. doi:10.1199/tab.0158.

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi:10.1016/j.plaphy.2013.02.001.

Sakuta, M. (2014). Diversity in plant red pigments: anthocyanins and betacyanins. *Plant Biotechnol. Rep.* 8, 37–48. doi:10.1007/s11816-013-0294-z.

Šarhanová, P., Pfanzelt, S., Brandt, R., Himmelbach, A., and Blattner, F. R. (2018). SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. *Ecol. Evol.* 8, 10817–10833. doi:10.1002/ece3.4533.

Sasaki, N., Nishizaki, Y., Ozeki, Y., and Miyahara, T. (2014). The Role of Acyl-Glucose in Anthocyanin Modifications. *Molecules* 19, 18747–18766. doi:10.3390/molecules191118747.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109, 365–371. doi:10.1007/s004120000089.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18, 233–234. doi:10.1038/72708.

Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi:10.1093/bioinformatics/bts094.

Schum, A., and Preil, W. (1998). “Induced Mutations in Ornamental Plants,” in *Somaclonal Variation and Induced Mutations in Crop Improvement* Current Plant Science and Biotechnology in Agriculture., eds. S. M. Jain, D. S. Brar, and B. S. Ahloowalia (Dordrecht: Springer Netherlands), 333–366. doi:10.1007/978-94-015-9125-6_17.

Sen, A., Ozturk, I., Yaycili, O., and Alikamanoglu, S. (2017). Drought Tolerance in Irradiated Wheat Mutants Studied by Genetic and Biochemical Markers. *J. Plant Growth Regul.* 36, 669–679. doi:10.1007/s00344-017-9668-8.

Sfeir, A., and Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem. Sci.* 40, 701–714. doi:10.1016/j.tibs.2015.08.006.

Shan, X., Li, Y., Yang, S., Gao, R., Zhou, L., Bao, T., et al. (2019). A functional homologue of Arabidopsis TTG1 from Freesia interacts with bHLH proteins to regulate anthocyanin and proanthocyanidin biosynthesis in both Freesia hybrida

and Arabidopsis thaliana. *Plant Physiol. Biochem.* 141, 60–72. doi:10.1016/j.plaphy.2019.05.015.

Shew, A. M., Nalley, L. L., Snell, H. A., Nayga, R. M., and Dixon, B. L. (2018). CRISPR versus GMOs: Public acceptance and valuation. *Glob. Food Secur.* 19, 71–80. doi:10.1016/j.gfs.2018.10.005.

Shinde, D. (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res.* 31, 974–980. doi:10.1093/nar/gkg178.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci.* 100, 15776–15781. doi:10.1073/pnas.2136655100.

Siebert, R., and Puchta, H. (2002). Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. *Plant Cell* 14, 1121–1131. doi:10.1105/tpc.001727.

Sikora, P., Chawade, A., Larsson, M., Olsson, J., and Olsson, O. (2011). Mutagenesis as a Tool in Plant Genetics, Functional Genomics, and Breeding. *Int. J. Plant Genomics* 2011, e314829. doi:https://doi.org/10.1155/2011/314829.

Silva, J. A. T. da, Shinoyama, H., Aida, R., Matsushita, Y., Raj, S. K., and Chen, F. (2013). Chrysanthemum Biotechnology: Quo vadis? *Crit. Rev. Plant Sci.* 32, 21–52. doi:10.1080/07352689.2012.696461.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi:10.1093/bioinformatics/btv351.

Slatnar, A., Mikulic-Petkovsek, M., Veberic, R., Stampar, F., and Schmitzer, V. (2013). Anthocyanin and chlorophyll content during poinsettia bract development. *Sci. Hortic.* 150, 142–145. doi:10.1016/j.scienta.2012.10.014.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., and Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26, 1134–1144. doi:10.1101/gr.196469.115.

Soliman, T. M. A., Lv, S., Yang, H., Hong, B., Ma, N., and Zhao, L. (2014). Isolation of flower color and shape mutations by gamma radiation of Chrysanthemum morifolium Ramat cv. Youka. *Euphytica* 199, 317–324. doi:10.1007/s10681-014-1127-z.

Sorensen, A.-M., Kröber, S., Unte, U. S., Huijser, P., Dekker, K., and Saedler, H. (2003). The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. *Plant J.* 33, 413–423. doi:10.1046/j.1365-313X.2003.01644.x.

Stadler, L. J. (1928a). Genetic Effects of X-Rays in Maize. *Proc. Natl. Acad. Sci. U. S. A.* 14, 69–75.

Stadler, L. J. (1928b). Mutations in Barley Induced by X-Rays and Radium. *Science* 68, 186–187. doi:10.1126/science.68.1756.186.

Stafford, H. A. (1994). Anthocyanins and betalains: evolution of the mutually exclusive pathways. *Plant Sci.* 101, 91–98. doi:10.1016/0168-9452(94)90244-5.

Stanley, L., and Yuan, Y.-W. (2019). Transcriptional Regulation of Carotenoid Biosynthesis in Plants: So Many Regulators, So Little Consensus. *Front. Plant Sci.* 10. doi:10.3389/fpls.2019.01017.

Steinmann, V. W., and Porter, J. M. (2002). Phylogenetic Relationships in Euphorbieae (Euphorbiaceae) Based on ITS and ndhF Sequence Data. *Ann. Mo. Bot. Gard.* 89, 453. doi:10.2307/3298591.

Stewart, R. N. (1960). Inheritance of Bract Color in Poinsettia. *J. Hered.* 51, 175–177. doi:10.1093/oxfordjournals.jhered.a106982.

Stewart, R. N., Asen, S., Massie, D. R., and Norris, K. H. (1979). The identification of poinsettia cultivars by HPLC analysis of their anthocyanin content. *Biochem. Syst. Ecol.* 7, 281–287. doi:10.1016/0305-1978(79)90005-X.

Stewart, R. N., Asen, S., Massie, D. R., and Norris, K. H. (1980). The anthocyanin and flavonol composition of three families of poinsettia color sports. *J. Hered.* 71, 140–142. doi:10.1093/oxfordjournals.jhered.a109329.

Stolc, V., Samanta, M. P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D. C., et al. (2005). Identification of transcribed sequences in Arabidopsis thaliana by using high-resolution genome tiling arrays. *Proc. Natl. Acad. Sci.* 102, 4453–4458. doi:10.1073/pnas.0408203102.

Sun, Y., Li, H., and Huang, J.-R. (2012). Arabidopsis TT19 Functions as a Carrier to Transport Anthocyanin from the Cytosol to Tonoplasts. *Mol. Plant* 5, 387–400. doi:10.1093/mp/ssr110.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* 9, 678–687. doi:10.1093/oxfordjournals.molbev.a040752.

Tanaka, A., Shikazono, N., and Hase, Y. (2010). Studies on Biological Effects of Ion Beams on Lethality, Molecular Nature of Mutation, Mutation Rate, and Spectrum of Mutation Phenotype for Mutation Breeding in Higher Plants. *J. Radiat. Res. (Tokyo)* 51, 223–233. doi:10.1269/jrr.09143.

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.

Taylor, J. M., Lopez, R. G., Currey, C. J., and Janick, J. (2011). The Poinsettia: History and Transformation. *Chron. Hortic.* 51, 23–28.

Taylor, L. P., and Grotewold, E. (2005). Flavonoids as developmental regulators. *Curr. Opin. Plant Biol.* 8, 317–323. doi:10.1016/j.pbi.2005.03.005.

Tian, L. (2015). Recent advances in understanding carotenoid-derived signaling molecules in regulating plant growth and development. *Front. Plant Sci.* 6. doi:10.3389/fpls.2015.00790.

Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., et al. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.* 7, 19. doi:10.1186/1471-2229-7-19.

Tohge, T., de Souza, L. P., and Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.* 68, 4013–4028. doi:10.1093/jxb/erx177.

Torregrosa, C., Cluzet, S., Fournier, J., Huguet, T., Gamas, P., Prospéri, J.-M., et al. (2004). Cytological, Genetic, and Molecular Analysis to Characterize Compatible and Incompatible Interactions Between Medicago truncatula and Colletotrichum trifolii. *Mol. Plant-Microbe Interactions*® 17, 909–920. doi:10.1094/MPMI.2004.17.8.909.

Trejo, L., Briones-Dumas, E., Gómez-Bermejo, R., and Olson, M. E. (2019). Molecular evidence for repeated recruitment of wild Christmas poinsettia (Euphorbia pulcherrima) into traditional horticulture in Mexico. *Genet. Resour. Crop Evol.* 66, 481–490. doi:10.1007/s10722-018-0727-1.

Trejo, L., Feria Arroyo, T. P., Olsen, K. M., Eguiarte, L. E., Arroyo, B., Gruhn, J. A., et al. (2012). Poinsettia's wild ancestor in the Mexican dry tropics: Historical, genetic, and environmental evidence. *Am. J. Bot.* 99, 1146–57. doi:10.3732/ajb.1200072.

Trejo, L., Rosell, J. A., and Olson, M. E. (2018). Nearly 200 years of sustained selection have not overcome the leaf area-stem size relationship in the poinsettia. *Evol. Appl.* 11, 1401–1411. doi:10.1111/eva.12634.

Trejo-Hernández, L., Olson-Zúnica, M. E., and Bye-Boettler, R. A. (2015). Datos históricos y diversidad genética de las nochebuenas (Euphorbia pulcherrima) del Distrito Federal, México. *Rev. Mex. Biodivers.* 86, 478–485. doi:10.1016/j.rmb.2015.04.033.

Ungaro, A., Pech, N., Martin, J.-F., McCairns, R. J. S., Mévy, J.-P., Chappaz, R., et al. (2017). Challenges and advances for transcriptome assembly in non-model species. *PLOS ONE* 12, e0185020. doi:10.1371/journal.pone.0185020.

USDA (2018). Floriculture Crops 2018 Summary. National Agricultural Statistics Service. Available at: https://www.nass.usda.gov/Publications/Todays_Reports/reports/floran19.pdf.

Valdés, A., Ibáñez, C., Simó, C., and García-Cañas, V. (2013). Recent transcriptomics advances and emerging applications in food science. *TrAC Trends Anal. Chem.* 52, 142–154. doi:10.1016/j.trac.2013.06.014.

van der Ploeg, R. (2018). FloraCulture International. *Int. Assoc. Hortic. Prod. IAHP*. Available at: https://www.floraculture.eu/digital-magazine/fci-november-december-2018/ [Accessed March 18, 2020].

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi:10.1016/j.tig.2014.07.001.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet. TIG* 34, 666–681. doi:10.1016/j.tig.2018.05.008.

Van Harten, A. M. (2002). “Mutation Breeding of Vegetatively Propagated Ornamentals,” in *Breeding For Ornamentals: Classical and Molecular Approaches*, ed. A. Vainstein (Dordrecht: Springer Netherlands), 105–127. doi:10.1007/978-94-017-0956-9_6.

van Nocker, S., and Ludwig, P. (2003). The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics* 4, 50. doi:10.1186/1471-2164-4-50.

Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., et al. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *R. Soc. Open Sci.* 3, 150565. doi:10.1098/rsos.150565.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science* 270, 484–487. doi:10.1126/science.270.5235.484.

Verstrepen, K. J., Jansen, A., Lewitter, F., and Fink, G. R. (2005). Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990. doi:10.1038/ng1618.

Verweij, W., Spelt, C., Di Sansebastiano, G.-P., Vermeer, J., Reale, L., Ferranti, F., et al. (2008). An H+ P-ATPase on the tonoplast determines vacuolar pH and flower colour. *Nat. Cell Biol.* 10, 1456–1462. doi:10.1038/ncb1805.

Vijay, N., Poelstra, J. W., Künstner, A., and Wolf, J. B. W. (2013). Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Mol. Ecol.* 22, 620–634. doi:10.1111/mec.12014.

Vik, N. I., Hvoslef-Eide, A. K., Gjerde, H., and Bakke, K. (2001). Stable transformation of pionsettia via electrophoresis. *Acta Hortic.*, 101–103. doi:10.17660/ActaHortic.2001.560.12.

Vilperte, V., Lucaciu, C. R., Halbwirth, H., Boehm, R., Rattei, T., and Debener, T. (2019). Hybrid de novo transcriptome assembly of poinsettia (Euphorbia pulcherrima Willd. Ex Klotsch) bracts. *BMC Genomics* 20, 900. doi:10.1186/s12864-019-6247-3.

Vimolmangkang, S., Han, Y., Wei, G., and Korban, S. S. (2013). An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol.* 13, 176. doi:10.1186/1471-2229-13-176.

Voshall, A., and Moriyama, E. N. (2018). “Next-Generation Transcriptome Assembly: Strategies and Performance Analysis,” in *Bioinformatics in the Era of Post Genomics and Big Data*, ed. I. Y. Abdurakhmonov (InTech). doi:10.5772/intechopen.73497.

Vu, G. T. H., Cao, H. X., Reiss, B., and Schubert, I. (2017). Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. *New Phytol.* 214, 1712–1721. doi:10.1111/nph.14490.

Wang, B., Kumar, V., Olson, A., and Ware, D. (2019a). Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Front. Genet.* 10. doi:10.3389/fgene.2019.00384.

Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., et al. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7, 11708. doi:10.1038/ncomms11708.

Wang, C.-S., Lo, K.-L., and Wang, A. Z. (2019b). Sodium azide mutagenesis generated diverse and broad spectrum blast resistance mutants in rice. *Euphytica* 215, 145. doi:10.1007/s10681-019-2468-4.

Wang, L., Tang, W., Hu, Y., Zhang, Y., Sun, J., Guo, X., et al. (2019c). A MYB/bHLH complex regulates tissue-specific anthocyanin biosynthesis in the inner pericarp of red-centered kiwifruit Actinidia chinensis cv. Hongyang. *Plant J.* 99, 359–378. doi:10.1111/tpj.14330.

Wang, M., Wang, P., Liang, F., Ye, Z., Li, J., Shen, C., et al. (2018). A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol.* 217, 163–178. doi:10.1111/nph.14762.

Wang, S., and Gribskov, M. (2017). Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis. *Bioinformatics* 33, 327–333. doi:10.1093/bioinformatics/btw625.

Wang, T., Wang, H., Cai, D., Gao, Y., Zhang, H., Wang, Y., et al. (2017). Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (Phyllostachys edulis). *Plant J. Cell Mol. Biol.* 91, 684–699. doi:10.1111/tpj.13597.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi:10.1038/nrg2484.

Ward, J. A., Ponnala, L., and Weber, C. A. (2012). Strategies for transcriptome analysis in nonmodel plants. *Am. J. Bot.* 99, 267–276. doi:10.3732/ajb.1100334.

Watanabe, K., Kobayashi, A., Endo, M., Sage-Ono, K., Toki, S., and Ono, M. (2017). CRISPR/Cas9-mediated mutagenesis of the dihydroflavonol-4-reductase-B (DFR-B) locus in the Japanese morning glory Ipomoea (Pharbitis) nil. *Sci. Rep.* 7, 10028. doi:10.1038/s41598-017-10715-1.

Watanabe, K., Oda-Yamamizo, C., Sage-Ono, K., Ohmiya, A., and Ono, M. (2018). Alteration of flower colour in Ipomoea nil through CRISPR/Cas9-mediated

mutagenesis of carotenoid cleavage dioxygenase 4. *Transgenic Res.* 27, 25–38. doi:10.1007/s11248-017-0051-0.

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 1155–1162. doi:10.1038/s41587-019-0217-9.

Werner, D. J., and Sink, K. C. (1977). Identification of poinsettia cultivars by electrophoretic analysis of proteins and peroxidases. *J. Hered.* 68, 35–40. doi:10.1093/oxfordjournals.jhered.a108769.

Whelan, S., and Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol. Biol. Evol.* 18, 691–699. doi:10.1093/oxfordjournals.molbev.a003851.

Winkel, B. S. J. (2004). Metabolic channeling in plants. *Annu. Rev. Plant Biol.* 55, 85–107. doi:10.1146/annurev.arplant.55.031903.141714.

Wu, C. A., Streisfeld, M. A., Nutter, L. I., and Cross, K. A. (2013). The Genetic Basis of a Rare Flower Color Polymorphism in Mimulus lewisii Provides Insight into the Repeatability of Evolution. *PLOS ONE* 8, e81173. doi:10.1371/journal.pone.0081173.

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., et al. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666. doi:10.1093/bioinformatics/btu077.

Xu, J., Hua, K., and Lang, Z. (2019). Genome editing for horticultural crop improvement. *Hortic. Res.* 6, 1–16. doi:10.1038/s41438-019-0196-5.

Xu, Z., Mai, Y., Liu, D., He, W., Lin, X., Xu, C., et al. (2020). Fast-Bonito: A Faster Basecaller for Nanopore Sequencing. *bioRxiv*, 2020.10.08.318535. doi:10.1101/2020.10.08.318535.

Yan, K., Wei, Q., Feng, R., Zhou, W., and Chen, F. (2017). Transcriptome analysis of Cinnamomum longepaniculatum by high-throughput sequencing. *Electron. J. Biotechnol.* 28, 58–66. doi:10.1016/j.ejbt.2017.05.006.

Yang, J., Zhang, J., Han, R., Zhang, F., Mao, A., Luo, J., et al. (2019). Target SSR-Seq: A Novel SSR Genotyping Technology Associate With Perfect SSRs in Genetic Analysis of Cucumber Varieties. *Front. Plant Sci.* 10. doi:10.3389/fpls.2019.00531.

Yao, P., Zhao, H., Luo, X., Gao, F., Li, C., Yao, H., et al. (2017). Fagopyrum tataricum FtWD40 Functions as a Positive Regulator of Anthocyanin Biosynthesis in Transgenic Tobacco. *J. Plant Growth Regul.* 36, 755–765. doi:10.1007/s00344-017-9678-6.

Yao, Y., and Kovalchuk, I. (2011). Abiotic stress leads to somatic and heritable changes in homologous recombination frequency, point mutation frequency and microsatellite stability in Arabidopsis plants. *Mutat. Res.* 707, 61–66. doi:10.1016/j.mrfmmm.2010.12.013.

Yasmin, K., Arulbalachandran, D., Soundarya, V., and Vanmathi, S. (2019). Effects of gamma radiation (γ) on biochemical and antioxidant properties in black gram (Vigna mungo L. Hepper). *Int. J. Radiat. Biol.* 95, 1135–1143. doi:10.1080/09553002.2019.1589022.

Yuan, H., Zhang, J., Nageswaran, D., and Li, L. (2015). Carotenoid metabolism and regulation in horticultural crops. *Hortic. Res.* 2, 1–11. doi:10.1038/hortres.2015.36.

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simão, F. A., Ioannidis, P., et al. (2017). OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 45, D744–D749. doi:10.1093/nar/gkw1119.

Zhang, B., Yang, X., Yang, C., Li, M., and Guo, Y. (2016). Exploiting the CRISPR/Cas9 System for Targeted Genome Mutagenesis in Petunia. *Sci. Rep.* 6. doi:10.1038/srep20315.

Zhang, H., Wang, L., Deroles, S., Bennett, R., and Davies, K. (2006a). [No title found]. *BMC Plant Biol.* 6, 29. doi:10.1186/1471-2229-6-29.

Zhang, L., He, L.-L., Fu, Q.-T., and Xu, Z.-F. (2013). Selection of reliable reference genes for gene expression studies in the biofuel plant Jatropha curcas using real-time quantitative PCR. *Int. J. Mol. Sci.* 14, 24338–54. doi:10.3390/ijms141224338.

Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H. (2006b). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat. Protoc.* 1, 641–646. doi:10.1038/nprot.2006.97.

Zhao, D., and Tao, J. (2015). Recent advances on the development and regulation of flower color in ornamental plants. *Front. Plant Sci.* 6. doi:10.3389/fpls.2015.00261.

Zhao, J. (2015). Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.* 20, 576–585. doi:10.1016/j.tplants.2015.06.007.

Zhao, J., and Dixon, R. A. (2010). The ‘ins’ and ‘outs’ of flavonoid transport. *Trends Plant Sci.* 15, 72–80. doi:10.1016/j.tplants.2009.11.006.

Zhao, M., Li, J., Zhu, L., Chang, P., Li, L., and Zhang, L. (2019). Identification and Characterization of MYB-bHLH-WD40 Regulatory Complex Members Controlling Anthocyanidin Biosynthesis in Blueberry Fruits Development. *Genes* 10, 496. doi:10.3390/genes10070496.

Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Luo, D., Li, X., and Hao, P. (2011). Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* 12, S2. doi:10.1186/1471-2105-12-S14-S2.

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE* 9, e78644. doi:10.1371/journal.pone.0078644.

Zimmermann, N. F. A., Ritz, C. M., and Hellwig, F. H. (2010). Further support for the phylogenetic relationships within Euphorbia L. (Euphorbiaceae) from nrITS and

trnL–trnF IGS sequence data. *Plant Syst. Evol.* 286, 39–58. doi:10.1007/s00606-010-0272-7.

Electronic appendix

Chapter 2 - Hybrid de novo transcriptome assembly of poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) bracts

Additional File 2. Annotation results of the poinsettia bract transcriptome against the TAIR10 database. (XLSX)

Additional File 4. BUSCO results for the *E. pulcherrima* and *E. pekinensis* transcriptomes. (XLSX)

Additional File 7. Differentially expressed transcripts for the comparisons between bract developmental stages in poinsettia. (XLSX)

Additional File 8. List of enriched GO terms for the comparisons between bract developmental stages in poinsettia. Comparison between the second stage (S2) of bract development against the third stage (S3). (XLSX)

Additional File 9. List of transcripts present in the differentially enriched pathways related to photosynthesis and phenylpropanoid metabolic processes. (XLSX)

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. Comparison between Christmas Feelings (red) and Christmas Feelings Pearl (white) varieties for the first stage (S1) of bract development. (XLSX)

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. Comparison between Christmas Feelings (red) and Christmas Feelings Pearl (white) varieties for the first stage (S1) of bract development. (XLSX)

Chapter 3 - A highly mutable GST is essential for bract colouration in *Euphorbia pulcherrima* Willd. Ex Klotsch

Additional File S1. Full-length sequence of the *Bract1* gene, CDS sequences of red and white poinsettia varieties, full sequence and truncated version of the *BRACT1* protein. (DOCX)

Additional File S5. Deducted protein sequences from the *Bract1* and 95 GSTs from *E. pulcherrima* GSTs, as well as anthocyanin-related GSTs from other species. (DOCX)

Additional File S7. Sequence alignment of *Bract1*, anthocyanin-related GSTs and orthologue GSTs from Euphorbiaceae species. (DOCX)

Chapter 4 - Development of a multiplex amplicon-sequencing assay to detect low-frequency mutations in poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) breeding programmes

Additional File S1. Number of reads containing either the *Bract1* wild-type or mutated alleles for each individual replicate of each sequencing library (XLSX).

Curriculum vitae

Personal information

Name	Vinicius Vilperte
Date of birth	08.02.1990
Place of birth	Florianópolis, Brazil
Nationality:	Brazilian
E-mail	vilperte@gmail.com
Linkedin	https://br.linkedin.com/in/vinicius-vilperte-83763a28
Research gate	https://www.researchgate.net/profile/Vinicius_Vilperte

Work experience

May 2020 – Present	Trainee in Sugar beet Genome Research at KWS SAAT SE & Co. KGaA (Einbeck, Germany)
May 2019 – April 2020	Research assistant at the Institute of Plant Genetics in the Department of Molecular Plant Breeding at the Leibniz Universität Hannover (Hannover, Germany)
May 2016 – May 2019	PhD candidate at Klemm + Sohn GmbH & Co. KG (Stuttgart, Germany)
January 2012 – June 2013	Exchange researcher at GenØk – Center for Biosafety (Tromsø, Norway)
August 2009 – December 2011	Junior researcher at Federal University of Santa Catarina (USFC) (Florianópolis, Brazil)

Academic education

May 2016 – May 2021	PhD student at the Leibniz Universität Hannover, Institute of Plant Genetics, Department of Molecular Plant Breeding (Prof. Thomas Debener) (Hannover, Germany)

March 2014 – February 2016	M.Sc. in Plant Genetic Resources at Federal University of Santa Catarina (UFSC), Post-Graduation Program in Plant Genetic Resource (Florianópolis, Brazil) Thesis: Comparative transcriptome and miRNome of transgenic maize varieties
March 2008 – December 2013	B.Sc. in Agronomy at Federal University of Santa Catarina (UFSC) (Florianópolis, Brazil) Thesis: Levels of DNA methylation and transcript accumulation in leaves of transgenic maize varieties

Publications

Vilperte V, Boehm R, & Debener T. (2021). Development of a multiplex amplicon-sequencing assay to detect low-frequency mutations in poinsettia (Euphorbia pulcherrima) breeding programmes. Plant Breeding 140: 497-507.

Vilperte V, Boehm R, & Debener T. (2021). A Highly Mutable GST is Essential for Bract Colouration in *Euphorbia Pulcherrima* Willd. Ex Klotsch. BMC Genomics, 22(1):208.

Vilperte V, Lucaciu CR, Halbwirth H, Boehm R, Rattei T, Debener T (2019). Hybrid de novo transcriptome assembly of poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch) bracts. BMC Genomics 20, 900.

Agapito-Tenfen SZ, **Vilperte V**, Traavik TI, Nodari RO (2018). Systematic miRNome profiling reveals differential microRNAs in transgenic maize metabolism. Envi. Sciences Europe 30:37.

Nitarska D, Stefanini C, Haselmair-Gosch C, Miosic S, Walliser B, Petkovsek MM, Regos I, Slatnar A, Debener T, Terefe-Ayana D, **Vilperte V**, Hadersdorfer J, Stich K, Halbwirth H (2018). 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 Biology 18:216.

Benevenuto RF, Agapito-Tenfen SZ, **Vilperte V**, Wikmark OG, van Rensburg PJ, Nodari RO (2017). Molecular responses of genetically modified maize to abiotic stresses as determined through proteomic and metabolomic analyses. PLoS ONE 12(2): e0173069.

Mesnage R, Agapito-Tenfen SZ, **Vilperte V**, Renney G, Ward M, Séralini GE, Nodari RO, Antoniou MN (2016). An integrated multi-omics analysis of the NK603 Roundup-tolerant GM maize reveals metabolism disturbances caused by the transformation process. Sci. Rep. 6: 37855.

Vilperte V, Agapito-Tenfen SZ, Wikmark OG, Nodari RO (2016). Levels of DNA methylation and transcript accumulation in leaves of transgenic maize varieties. Envi. Sciences Europe 28: 29.

Donazzolo J, Ornellas TS, Bizzocchi L, **Vilperte V**, Nodari RO (2015). O armazenamento refrigerado prolonga a viabilidade de sementes de goiabeira-serrana. Revista Brasileira de Fruticultura 37: 748.

Ambrosini VG, Voges JG, Benevenuto RF, **Vilperte V**, Abreu Silveira M, Brunetto G, Ogliari JB (2015). Single-head broccoli response to nitrogen application. Científica (Jaboticabal) 43: 84-92.

Agapito-Tenfen SZ, **Vilperte V**, Benevenuto RF, Rover CM, Traavik TI, Nodari RO (2014). Effect of stacking insecticidal cry and herbicide tolerance epsps transgenes on transgenic maize proteome. BMC Plant Biology 14: 346.

Klabunde GHF, Olkoski D, **Vilperte V**, Zucchi MI, Nodari RO (2014). Characterization of 10 new nuclear microsatellite markers in Acca sellowiana (Myrtaceae). Applications in Plant Science 2: 1400020.

Conference poster contributions

Vilperte V, Boehm R, Debener T. (2020) Glutathione S-transferase as a potential marker for mutation breeding in poinsettia (*Euphorbia pulcherrima* Willd. ex Klotsch). In: Annals of the International Symposium of the Society for Plant Breeding e.V. (GPZ), Tulln – Austria.

Vilperte V, Lucaciu CR, Boehm R, Rattei T, Debener T. (2018) Hybrid de novo transcriptome assembly of Poinsettia (*Euphorbia pulcherrima* Willd. Ex Klotsch). In: Annals of the 12th Congress of the International Plant Molecular Biology, Montpellier – France.

Nitarska D, Haselmair-Gosch C, **Vilperte V**, Terefe-Ayana D, Debener T, Stich K, Halbwirth H. (2017) Evaluation of poinsettia (Euphorbia pulcherrima) for breeding

blue strategies identified genes encoding inactive F3. In: Annals of the 9th International workshop on Anthocyanins, 2017, Auckland. p. 68.

Agapito-Tenfen SZ, **Vilperte V**, Traavik TI, Nodari RO (2015). miRNome profiling of Bt and EPSPS- expressing transgenic maize. In: Proceedings of the 2nd EFSA Scientific Conference, Milan – Italy, 14-16 October 2015. EFSA Journal 2015, 13(10):s1310, 137 pp.

Benevenuto RF, Agapito-Tenfen SZ, **Vilperte V**, Nodari RO (2014). Comparative proteomic analysis of transgenic maize (NK603) and its non- transgenic counterpart in response to drought stress. In: Annals of the the 9th International Conference on Genomics, Shenzen.

Vilperte V, Donazzolo J, Klabunde GHF, Nodari RO (2014). Urban gardens as gene pools for Acca sellowiana (O.Berg) Burret diversity conservation. In: XLIII Congreso Argentino de Genética, 2014, San Carlos de Bariloche. Journal of Basic & Applied Genetics, XXV:272.

Klabunde GHF, **Vilperte V**, Stefenon VM, Nodari RO (2012). Phylogeographical analysis within populations of Araucaria angustifolia Bert. O. Kuntze in its natural distribution range. In: Annals of the 4th EMBO Meeting - Advancing the life sciences.

Vilperte V, Olkoski D, Klabunde GHF, Petry VS, Nodari RO (2011). Analysis of genetic diversity in a natural population suggest long distance seed dispersion of Acca sellowiana (Berg.) Burret. In: Annals of the 57th Congresso Brasileiro de Genética, 2011, Águas de Lindóia - SP.

Vilperte V, Rosa AM, Ambrosini VG, Benevenuto RF, Agapito-Tenfen SZ, Elias HT, Nodari RO (2010). Differential physiological responses of transgenic maize variety compared to its non- transgenic counterpart under abiotic stress during germination. In: Annals of the 11th International Symposium on the Biosafety of Genetically Modified Organisms (ISBGMO), p.188 – 190.

Patents

Vilperte V, Debener T, Boehm R, Dohm A, Von Tubeuf G, Sander U (2020). Method of generating plants having white foliage (European Patent No - WO2020244759). European Patent Office.

Acknowledgement

First, I would like to thank Prof. Dr. Thomas Debener not only for the outstanding supervision but also for allowing me to develop myself personally and professionally and for helping me through the challenges that a doctorate entail. Thank you for the support, the countless discussions and trust during all these years.

I would also like to thank Prof. Helge Küster and Prof. Traud Winkelmann for taking part in the examination and for providing valuable feedback.

Next, I would like to thank all the staff of the Department of Molecular Plant Breeding for a fantastic time and for welcoming me to the team. Thanks to you, the work in the lab was so enjoyable, the coffee breaks were always filled with cake and laugh, and one could always count on your support. In particular, I would like to thank Ingrid Robotta for the constant help and commitment in the lab and Julia Schröter for the immense help and for the patience to answer all my questions.

I would especially like to mention my fellow doctoral students Ina Menz, Annette Bartkiewicz, Enzo Neu, Friederike Chilla, Helena Domes, Juliane Geike and Jannis Straube. From the first day I arrived in Hannover, you welcomed and supported me in every way possible. Thanks for all the laughing, the joking, the singing, and for the support during the ups and downs of this journey. You made this time extraordinary and unforgettable. Thank you very much!

Another big thank you goes to everyone from the FlowerPower team for all the fruitful meetings and discussions. A special thank you to my FlowerPower doctorate colleagues Martina Kolarek, Daria Nitarska and Carmen Stefanini. Your company and friendship in all the meetings, presentations and all the places we visited together made this journey even more special.

Finally, I would like to thank all my friends that help me through this time, and of course my family. Thanks mom and dad for always supporting and encouraging me to achieve my goals. Last but not least, my heartful thanks goes to Kai Krüger, who was always there to listen and support me in every step of my professional and personal growth.