Annals of Botany

Original Article

An approach to determining anthocyanin synthesis enzyme gene expression in an evolutionary context: an example from *Erica plukenetii*.

N C Le Maitre[1,a], M D Pirie[2,3] and D U Bellstedt*[1]

1) Biochemistry Department, Stellenbosch University, Private Bag X1, Matieland, 7600, South Africa. 2.) Johannes Gutenberg-Universität, Mainz, 55099 Mainz, Germany. 3) University Museum, University of Bergen, Postboks 7800, N-5020 Bergen, Norway. a) Current address: University of the Free State, Bloemfontein, PO Box 339, 9300, South Africa.

Running title: Anthocyanin synthesis in *Erica plukenetii*

Corresponding author: dub@sun.ac.za

Summary

- Background and Aims:

Floral colour in angiosperms can be controlled by variations in the expression of the genes of the anthocyanin pathway. Floral colour shifts influence pollinator specificity. Multiple shifts in floral colour occurred in the diversification of the genus *Erica* (Ericaceae), from plesiomorphic pink to e.g. red or white flowers. Anthocyanin gene expression variation and its effects on floral colour in the red-, pink- and white flowered *Erica plukenetii* species complex was investigated.

- Methods:

Next Generation Sequencing, Reverse Transcriptase PCR, and Real Time Reverse Transcriptase quantitative PCR were used to quantify anthocyanin gene expression.

- Key results:

Non-homologous mutations causing loss of expression of single genes were found, indicating that the cause was likely to be mutations in transcription factor binding sites upstream of the 5'-UTR of the genes and this was confirmed by sequencing.

- Conclusions:

Independent evolution and subsequent loss of expression of anthocyanin genes may have influenced diversification in the *Erica plukenetii* species complex. The approach developed here should find more general application in studies on the role of floral colour shifts in diversification.

- Key words: Anthocyanin biosynthesis, gene expression, *Erica*, evolution, floral colour, RT-qPCR

Introduction

Flower colour is an important factor in pollinator specificity in angiosperms, and shifts in colour are important in pollinator-mediated speciation (Rausher 2008; Carlson and Holsinger 2013). The anthocyanin biosynthesis pathway (Grotewold 2006) forms the anthocyanin pigments that colour plant tissues in general and are a major determinant of floral colour. The pathway consists of at least six enzymes, in order: chalcone synthase (CHS), chalcone isomerase (CHI), flavanone hydroxylase (F3H), dihydroflavanol-4-reductase (DFR), anthocyanin synthase (ANS) and UDP 3-O-glucosyltransferase (UDP-GST) to produce pelargonidin. The addition of flavanoid 3' hydroxylase (F3'H) or flavanoid 3' 5' hydroxylase (F3'5'H) after F3H produces cyanidin or delphinidin respectively. The transcription of the enzymes' genes is collectively regulated by a complex of three transcription factors, Ip-WDR, Ip-MYB and Ip-bHLH in *Ipomoea purpurea* (Zhu *et al.* 2015), that bind to the MYB recognition element (MRE) and the bHLH recognition element (BRE) site upstream of the 5' UTR of each of the genes. Shifts in floral colour occur as a result of functional mutations in the transcription factor genes, loss of function mutations in the genes of the anthocyanin pathway themselves or mutations to the upstream transcription factor binding sites (Wessinger and Rausher 2012).

To investigate the general process of shifts in flower colour associated with changes in anthocyanin biosynthesis, it is important to examine multiple such shifts in a biologically comparable system. To that end, we have been investigating a group of flowering plants in which shifts in flower colour have been frequent during the origins of a vast species diversity: the genus *Erica* (Ericaceae). *Erica* is the most species-rich genus in the Cape Floristic Region (CFR) with ca. 690 species (Oliver 1989, 1999; Linder 2003), and hence a model group for investigating the factors driving evolutionary diversification.

Phylogenetic hypotheses for *Erica* based on nuclear and plastid DNA sequence markers and up to 60% of the known species diversity (Pirie *et al.* 2011, 2016; Mugrabi De Kuppler *et al.* 2015) have been used to infer the age and diversification rate of the clade, as well as patterns of character evolution. The results showed that speciation in the most species-rich Cape clade was rapid (Pirie *et al.* 2016), and that multiple shifts in floral morphology had occurred (Pirie *et al.* 2011; Le Maitre 2017). These shifts led to changes between wind-, insect- and bird pollination syndromes and hence restriction of gene flow between populations that may have played a role in driving speciation (Pirie *et al.* 2011). Along with the shifts in overall pollination syndrome, frequent changes in floral colour between the plesiomorphic pink (predominant in the older European lineages (Mugrabi De Kuppler *et al.* 2015)) and derived colours ranging from white (observed in species across the genus) through green, yellow, and red (restricted to those of the diverse Cape clade) were inferred (Le Maitre 2017).

Anthocyanins are the primary determinants of floral colour in *Erica* species (Crowden and Jarman 1976). Only cyanidin is present in *Erica plukenetii*, giving it its red/pink colouration. The sequences of the anthocyanin genes and their *trans*-acting regulatory factors in *Erica* have not been investigated before. We chose to focus initially on one species, *E. plukenetii* because it shows floral colour polymorphism both within and between populations, has subspecies that are constitutively red or white, and it is hypothesised that the shifts in floral colour may influence pollinator shifts between bird- and insect pollination (van der Niet *et al.* 2014).

To study the molecular changes underlying these shifts in floral colour, first the anthocyanin genes in *E. plukenetii* had to be characterised. To this end, a Next Generation Sequencing (NGS) approach was used to sequence the genome, in which the anthocyanin genes could be found. Primers could be designed on the exons of the genes for Real Time Reverse Transcriptase quantitative PCR (RT-qPCR) determinations of the expression of the genes of

the anthocyanin pathway in differently coloured flowers. If the expression of the genes was identical in red- and white flowered plants, it would point to a loss of function mutation in a gene of the pathway. If the expression of all the genes of the pathway was knocked down in the white flowered plants, it would point to a transcription factor mutation. If the expression of a single gene in the white flowered plants was knocked down, it would point to a mutation in the transcription factor binding sites regulating that specific gene. The establishment of which mechanism underlies the floral colour shift in the *E. plukenetii* complex may give insights into the mechanism underlying the other independent floral colour shifts in the genus.

Materials and Methods:

DNA was extracted from a red flowered *E. plukenetii*, collected from Table Mountain, Cape Town, South Africa, by Plant Research International at Wageningen University, the Netherlands, using their in-house CTAB extraction protocol. The Illumina library was prepared and one Illumina HiSeq2500 paired end run with 250 bp reads and two Illumina MiSeq paired end runs with 2x300 bp reads were performed and the data was assembled using CLC Bio. A local BLAST database was created in Geneious R9 (Kearse *et al.* 2012) from the contiguous sequences. Exon sequences are highly conserved and are therefore ideal for finding homologous sequences via local tBLASTx searches. The presence of non-conserved intron sequences in complete gene sequences or the concatenation of exon sequences in mRNA transcripts of genes, lowers the overall sequence homology and makes finding genes much less likely, especially when genes are potentially spanned across more than one contiguous sequence. Therefore, it was necessary to use sequences of the anthocyanin pathway genes and their transcription factors in which exons were annotated and thus could be used individually in tBLASTx searches of the local BLAST database. For each

gene, the fully annotated sequence from the most closely related species that could be identified on Genbank was used. In some cases these were represented by a species in the same family: Ericaceae (*Rhododendron simsii*); in others we were obliged to compare across the wider order Ericales (*Camellia sinensis*, *Camellia nitidissima*, *Diospyros kaki*, *Vaccinium corymbosum*, and *Actinidia chinensi*)*;* or, in the case of ANS, to a distantly related angiosperm clade, Rosids (*Vitis vinifera*; APG IV (Chase *et al.* 2016)). Transcription factor binding sites were found using the Motif Finder in Geneious R9 and a Perl script (Zhu *et al.* 2015). Primer pairs were designed to amplify overlapping regions from the upstream regulatory regions of each gene to the end of the last exon based on the aligned exons from the tBLASTx searches. Where contiguous sequences had gaps, or a gene spanned multiple contiguous sequences, a chromosome walking approach was used to complete the sequences. DNA was directly amplified (Bellstedt *et al.* 2010) from the leaves of the same sample of red flowered *E. plukenetii* ssp. *plukenetii* that was used for the Illumina sequencing. PCR amplifications were carried out using the Kapa Biosystems 3G Plant PCR kit. Each 25 µl PCR mix contained 12.5 µl Buffer, 2 µl 25 mM $MgCl_2$, 0.75 µl 20 mM Forward and 0.75 µl 20 mM Reverse Primers (Table 1), 0.2 µl 2.5 U/µl 3G PlantTaq, 0.2 µl 5% DMSO, 1 µl extracted DNA and 8.6 µl milliQ H_2O in an Applied Biosystems Veriti PCR thermal cycler. The PCR program was 2 minutes at 95°C; followed by 35 cycles of 95°C for 30 seconds, annealing at the temperature specified in Table 1 for 30 seconds, and extension at 72°C for 30 seconds; followed by a final extension step of 72°C for 5 minutes. Sanger sequencing reactions were performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fischer Scientific) in an Applied Biosystems PCR thermal cycler using the STeP sequencing protocol (Platt *et al.* 2007). Each 10 µl sequencing reaction contained 5 µl Sequencing Buffer, 1 µl 0.8 mM Forward Primer or Reverse Primer (Table 1), 1 µl Big Dye, 1 µl of PCR product and 2.5 µl milliQ water. Sequencing electrophoresis was performed at the Central Analytical

Facility at Stellenbosch University. Primers for RT-qPCR (Table 2) were designed on the sequences of the exons of the anthocyanin pathway genes. *Paratubulin 1* (*Ep-PBT1*) was identified as a candidate reference gene, as it has been shown to be expressed stably in *C. sinensis* (Hao *et al.* 2014) and its expression was tested for consistency in floral tissue using the same methods as Hao *et al.* (2014). Total RNA was extracted from 50 mg of floral tissue using the Qiagen RNeasy Power Plant Kit with the Qiagen RNase-Free DNase Set from corollas of eight different lengths (Supplementary Figure 1) of the collected red-, pink- and white flowered *E. plukenetii* ssp. *plukenetii* samples from a population on Du Toits Kloof Pass (DP), red- and white flowered *E. plukenetii* ssp. *plukenetii* from Franschhoek Pass (FP) (Supplementary Figure 2), as well as from corollas of three different lengths of white flowered *E. plukenetii* ssp. *breviflora* in the Witzenberg Valley (WV), South Africa. Only three corolla lengths of *E. plukenetii* ssp. *breviflora* were collected as only three growth points were easy to differentiate, due to its reduced corolla length. RNA quality was verified using gel electrophoresis. RNA was quantified using a Nanodrop ND-1000 spectrophotometer, and the 260/280 absorbance ratios were measured to check RNA quality.

In the samples collected from the DP population, expression was tested using Reverse Transcriptase PCR (RT-PCR) and agarose gel electrophoresis to test the feasibility of the approach. The population was later burnt in a fire, precluding the collection of further samples for RT-qPCR analysis. The Kapa Plant 3G Plant PCR kit was used. Each RT PCR mix contained 10 µl PCR Buffer; 0.4 µl 10 mM Forward and 0.4 µl 10 mM Reverse Primer (see Table 2); 0.4 µl 50X Reverse Transcriptase; 3 µl extracted RNA 5.8 µL milliQ H_2O to make up a total of 20 µl. RT PCR was performed in a Applied Biosystems Veriti PCR machine, with a single pre-incubation step of 42°C for 5 minutes then 95°C for 3 minutes, followed by 45 cycles of 95°C for 3 seconds, 60°C for 20 seconds and 72°C for 2 seconds. PCR products were separated and visualised on a 1.5% agarose gel.

In the samples collected from the FP population, the expression of the genes of the anthocyanin biosynthesis pathway enzymes was determined using Real Time Reverse Transcriptase quantitative PCR (RT-qPCR) with three biological replicates per corolla growth point, using the Kapa Probe Fast One Step RT-qPCR kit. Each multiplex RT-qPCR mix contained 10 µl Buffer; 2 or 3 sets of: 0.4 µl 10 mM Forward and 0.4 µl 10 mM Reverse Primer (see Table 1), 0.4 µl 10 mM probe (see Table 2); 0.4 µl 10 mM PBT1-1,416F, 0.4 µl 10 mM PBT1-1,564R and 0.4 µl 10 mM PBT1-1500Probe; 0.2 µl 2.5 U/µl 3G Plant Taq; 1 µl extracted RNA and sufficient milliQ H_2O to make up a total of 20 µl. Up to four reactions were multiplexed together in a single reaction. RT-qPCR was performed in a Roche LightCycler 96, with a single pre-incubation step of 42°C for 5 minutes then 95°C for 3 minutes, followed by 45 cycles of 95°C for 3 seconds, 60°C for 20 seconds and 72°C for 2 seconds.

In the *E. plukenetii* ssp. *breviflora* WV population, only some of the probes bound to their respective templates successfully, so quantifications were performed using Kapa SYBR Fast Universal Kit. Each RT-qPCR mix contained 10 µl PCR Buffer; 0.4 µl 10 mM Forward and 0.4 µl 10 mM Reverse Primer (see Table 2); 0.4 µl 50X Reverse Transcriptase; 3 µl extracted RNA 5.8 µL milliQ H_2O to make up a total of 20 µl. RT-qPCR was performed in a Roche LightCycler 96, with two preincubation steps of 61°C for 5 minutes, then 95°C for 30 seconds, followed by 50 cycles of 95°C for 3 seconds, 58°C for 10 seconds and 72°C for 2 seconds. A melting step followed with 95°C for 10 seconds, 65°C for 1 minute and 95°C for 1 second, verifying the presence of a single amplicon per reaction.

No-template and no-reverse transcriptase controls were included with each batch of samples analysed in all assays. Expression of the *trans*-acting transcription factor genes was determined using conventional RT-PCR and expression levels were assessed using gel electrophoresis. For the RT-qPCRs, expression of the genes of the anthocyanin biosynthesis

pathway enzymes was measured relative to background fluorescence and was normalised to the expression of the reference gene, *Ep-PBT1*, using the Roche LightCycler 96 software. The RT-qPCR assays were optimised to ensure that all the genes had similar amplification efficiency, slope and Cq values. Melting curve analysis was used to check for the presence of a single amplicon with the SYBR Green RT-qPCRs.

Results:

The assembled Illumina NGS data of the red flowered *E. plukenetii* contained 602 million reads, totalling more than eight billion bases. These reads were assembled using CLC Bio into 400 000 contiguous sequences. Actual coverage of the genome was around 20X. Of the assembled 4 443 321 contiguous sequences, only 3.75% were larger than 1000 bp, with an N50 of 1 673.

All eight genes of the anthocyanin biosynthesis pathway enzymes and their *trans*-acting transcription factor genes (Grotewold 2006; Zhu *et al.* 2015), were found in the NGS data using the tBLASTx approach. Exons of a particular gene (from Genbank) were aligned to the contiguous sequence(s), and the annotations were transferred to the contiguous sequence(s). If the aligned exons spanned multiple contiguous sequences, the contiguous sequences were concatenated. Start codons were identified and the stop codon at the end of the final exon was assumed to be the end of the gene. The *cis* BRE and MRE motifs upstream of the transcription start site were found in all of the contiguous sequences, except for that of *Ep-CHI*, where the region upstream of the start of the first exon of the gene was not present in the NGS data. Sequence identity in the aligned exonic regions approached 90%. Chromosome walking using Sanger sequencing was successfully used to obtain any missing sequence data in or between contiguous sequences (Le Maitre 2017). The sequences of the anthocyanin biosynthesis pathway genes and their *trans*-acting transcription factor genes

from *E. plukenetii* ssp. *plukenetii* were deposited on Genbank, with accessions: MG948576 (*Ep-ANS*), MG948577 (*Ep- bHLH*), MG948578 (*Ep-CHI*), MG948579 (*Ep-CHS*), MG948580 (*Ep-DFR*), MG948582 (*Ep-F3'H*), MG948583 (*Ep-F3H*), MG948584 (*Ep-R2R3-MYB*), MG948585 (*Ep-UDP-GST*), MG948586 (*Ep-WD40*) and MG948587 (*Ep-PTB1*).

RNA was successfully isolated from the samples collected from DP, FP and WV populations. RNA quality was verified using gel electrophoresis. Contamination with gDNA was minimal. *Ep-PBT1* was found to be stably expressed in floral tissues and was therefore used to normalise the expression of the genes. The three *trans*-acting genes that regulate the expression of the anthocyanin genes *Ep-R2R3 MYB*, *Ep-bHLH*, and *Ep-WDR* were found to be expressed at comparable levels in all samples from all three populations.

The anthocyanin genes were expressed in a coordinated manner. Expression levels increased with increasing corolla growth point and then declined as the flowers became mature. Red- and pink flowered *E. plukenetii* ssp. *plukenetii* from DP (Figure 1) and FP (Figure 2) expressed all the anthocyanin genes. In the white flowered *E. plukenetii* ssp. *plukenetii* from DP, *chalcone synthase* (*Ep-CHS*) (Figure 1) was not expressed and in the white flowered *E. plukenetii* ssp. *plukenetii* from FP (Figure 3), *anthocyanin synthase* (*Ep-ANS*) was not expressed above the background fluorescence. In *E. plukenetii* ssp. *breviflora* from WV, *UDP-3-0-glucosyltransferase* (*Ep-UDP-GST*) was not expressed above the background fluorescence and the *dihydroflavanol-4-reductase* (*Ep-DFR*) was very low relative to both *Ep-PTB1* and the other genes (Figure 4), but it was expressed above the background fluorescence at growth point 2. Variation in individual gene expression levels in *E. plukenetii* ssp. *breviflora* was greater than in *E. plukenetii* ssp. *plukenetii*.

The 5' upstream regions of the *Ep-CHS* gene from white flowered *E. plukenetii* ssp. *plukenetii* (DP), the *Ep-ANS* gene from white flowered *E. plukenetii* ssp. *plukenetii* (FP) and

the *Ep-DFR* and *Ep-UDP-GST* genes from white flowered *E. plukenetii* ssp. *breviflora* (WV) were sequenced to find any mutations that could account for the lack of expression observed. A deletion was found in the BRE motif of *Ep-CHS* in *E. plukenetii* ssp. *plukenetii* (DP), a CC to AA mutation was found in the MRE motif of *Ep-ANS* in *E. plukenetii* ssp. *plukenetii* (FP) and a C to A mutation was found in the MRE motif of *Ep-UDP-GST* of *E. plukenetii* ssp. *breviflora* (WV) (Figure 5). No mutations were found in the BRE or MRE motifs of the *Ep-DFR* gene of *E. plukenetii* ssp. *breviflora* (WV).

Discussion

The NGS sequencing approach was successful but contiguous sequence lengths were typically too short to contain entire genes and exons often had to be located on separate contiguous sequences and concatenated. Using flow cytometry, the *Erica plukenetii* genome has been shown to have a 2C value of 0.91, which equates to around 900 Mb (van Der Niet, pers. comm.). This fits well with the known genome sizes from related genera of Ericaceae that are c. 695 Mb (*Rhododendron delavayi*; Ericoideae [the same subfamily as *Erica*] (Zhang *et al.* 2017)) and c. 420 Mb (*Vaccinium macrocarpum*; more distantly related Vaccinioideae (Polashock *et al.* 2014)). There are no known polyploids of *Erica* species, with known counts from wild species restricted to 2N=24 in most species (Nelson and Oliver 2005). Our coverage was not sufficient for a complete genome assembly, and hence for a precise estimate of genome size in *Erica*, but this was to be expected given the number of reads we obtained and a genome size similar to, or larger than, that known from other Ericaceae.

All of the eight genes of the anthocyanin biosynthesis pathway enzymes and their three *trans*-acting transcription factor genes were found in the NGS data generated from *E. plukenetii* ssp. *plukenetii*. The tBLASTx approach to locating the genes of interest in the NGS data will

have broader application to other studies where little to no genetic data is available on the genes of interest.

The exons of the eight anthocyanin genes and the transcription factor genes are highly similar to the exons (exon sequence homology approached 90%) of those same genes in the most closely related species for which data was available, from the same family: *Rhododendron simsii* (Ericaceae); same order: *Camellia sinensis*, *Camellia nitidissima*, *Diospyros kaki*, *Vaccinium corymbosum*, *Actinidia chinensis* (Order: Ericales); and *Vitis vinifera* (Order: Rosids), indicating that their functions are both conserved and important.

The clustering of the expression levels of all the genes at each corolla length indicates the expression of genes coding for enzymes of the anthocyanin biosynthesis pathway is regulated by a common method of transcriptional regulation, with expression of all the genes being roughly equal at a given time. This is notably similar to the results of Zhu *et al.* (2015). The loss of pigmentation in the DP, FP and WV populations appears to be linked to mutations in the *cis* motifs that are involved in the regulation of expression. Alternative explanations, such as mutations in the transcription factors themselves (that would lead to global knockdown of anthocyanin gene expression) or a loss of function mutation in one of the genes of the pathway were not supported. While mutations in the *cis* motifs to which the *trans*-acting transcription factors bind are not the most common cause of the loss of red pigmentation, they have been observed in 12% of studies documented by Streisfeld and Rausher (2011), and site directed mutagenesis of the BRE and MRE motifs in *I. purpurea* was found to significantly reduce expression of anthocyanin pathway genes (Zhu *et al.* 2015).

The existence of the BRE and MRE motifs upstream of the anthocyanin genes, the loss of gene expression associated with mutations in the motifs and the coordinated expression of the genes of the anthocyanin biosynthesis pathway would seem to indicate that the mechanism of

control of anthocyanin biosynthesis in *E. plukenetii* is the same as in *Ipomoea* (Zhu *et al.* 2015), *Mimulus* (Yuan *et al.* 2014) and *Arabidopsis* (Koes *et al.* 2005; Grotewold 2006).

Point mutations in the *cis* motifs of the *Ep-CHS*, *Ep-ANS* and *Ep-UDP-GST* genes could disrupt the binding of the *trans*-acting Ep-MYB- Ep-bHLH- Ep-WDR complex. The disruption of the binding of the *trans*-acting factors to the motifs could lead to reduction or total knockdown of the expression of the *Ep-CHS, Ep*-ANS and *Ep-UDP-GST* genes, as has been observed in *Ipomoea purpurea* (Zhu *et al.* 2015). Consequently, the Ep-CHS, Ep-ANS and Ep-UDP-GST enzymes would not be produced.

If Ep-CHS is not produced, chalcone is not synthesised (Wessinger and Rausher 2012; Zhu *et al.* 2015). The loss of chalcone would have implications for plant fitness, as it is not only a precursor for the anthocyanin pathway but also forms part of the immune system (Dao *et al.* 2011; Rozmer and Perjési 2016) and plays a role in reducing the effects of heat stress and UV damage (Coberly and Rausher 2003). If Ep-ANS and Ep-UDP-GST are not produced, the direct consequences are less obvious, as the intermediates are committed to the pathway (Wessinger and Rausher 2012). There would however be implications for the plant beyond the loss of colour, as anthocyanins play multiple roles in plant tissues including protection of photosynthetic molecules and chelation of photosynthetic by-products (Gould *et al.* 1995; Smillie and Hetherington 1999; Landi *et al.* 2015), and protection of other photolabile compounds in leaves (Gould *et al.* 2010).

The mutations associated with the observed colour shifts appear to have independent origins as they differ between both the white flowered populations of *E. plukenetii* ssp. *plukenetii* and also from the population of *E. plukenetii* ssp. *breviflora*.

Neither the heritability nor any fitness effects of these mutations have been tested, nor has the incidence of these mutations within the sampled populations been quantified. However, they

appear to be tolerable across generations, as white flowered plants of various ages occur in the populations. Hence the same mutations may occur in other closely related white flowered *E. plukenetii* populations, whilst similar mutations could be independently derived in more distantly related ones.

From a phylogeny of *E. plukenetii* (van der Niet *et al.* 2014), it appears that the plesiomorphic form, similar to that of *E. plukenetii* ssp. *plukenetii*, has pink or red, medium length, unscented corollas; and an upright growth form with sturdy branches suitable for perching on by birds. *E. plukenetii* ssp. *breviflora* by contrast has white, short, scented corollas attractive to its moth pollinators, with flowers presented on long thin branches that discourage bird perching (van der Niet *et al.* 2014). Although the latter differences in habit represent part of the overall pollination syndrome, they may primarily reflect adaptations to the differing habitats of the subspecies: mountains in *E. plukenetii* ssp. *plukenetii* and sandy plains in *E. plukenetii* ssp. *breviflora.* The major pollinator of the *E. plukenetii* species complex, the Orange Breasted Sunbird, *Anthobaphes violacea*, is present in both habitats. In this scenario, the shift to moth pollinaton would not be related to the distribution of the pollinator, but instead a consequence of adaptation to the differing environment, followed by reinforcement through further pollinator-specific adaptations, including reduction in flower size and production of scent (van der Niet *et al.* 2014). It is not clear at what stage in this process white flower colour would have become fixed.

The loss of red flower colour is therefore just one of a suite of adaptations contributing to the shift from bird to insect pollination in *E. plukenetii*, and potentially it was neither the first nor the most critical factor involved in the process. However, the importance of flower colour changes in pollinator mediated diversifications has been shown in a range of plant groups (Bradshaw and Schemske 2003; Hoballah *et al.* 2007; Wessinger and Rausher 2012). Amongst the wide variation in flower colour in Cape *Erica*, many taxa are either white

flowered, or show white flowered forms, including within species complexes such as in the *E. abietina*/*E. viscaria* clade (e.g. white flowered *E. viscaria* ssp. *pendula*; white forms of *E. regia* ssp. *regia* and of *E. vestita* (Pirie *et al.* 2017)). In general, the kind of complex scenario apparent for *E. plukenetii*, involving an interplay of adaptations to differing ecological and pollination niches, may drive elevated rates of speciation in *Erica* and other species-rich genera in the CFR (Linder 2003; Pirie *et al.* 2011, 2016). Importantly, the simple general mechanism underlying the change from pink to white flowers may explain both the regular occurrence of white flower colour in *Erica* species and its contribution to the prevalence of pollinator shifts in the radiation of the *Erica* clade.

This is the first study to identify the genes and their putative *trans*-regulatory factors, of the anthocyanin biosynthesis pathway in a species of *Erica,* and to propose a mechanism for floral colour changes observed in *Erica*. This study represents an important step towards understanding specific mechanisms that may have contributed to the rapid speciation inferred in *Erica,* particularly in the Cape clade. The techniques and approaches developed will allow testing of the mechanisms of floral colour shifts in other *Erica* species complexes or species pairs where floral colour shifts from the plesiomorphic pink to red or white flowered species have occurred and identify the commonalities between them.

Funding

This work was supported by the South African National Research Foundation [grant number 98867]

Acknowledgements

The authors would like to thank South African National Parks and Cape Nature for allowing sample collection and the University of Stellenbosch, Department of Biochemistry for the facilities.

Author contributions

N Le Maitre collected samples, performed all the molecular analyses except the NGS, did all the data analysis and wrote the first draft of the paper. D Bellstedt collected samples, conceptualised the project and edited the drafts. M Pirie provided substantial editorial input and support.

Literature cited

Bellstedt DU, Pirie MD, Visser JC, de Villiers MJ, Gehrke B. **2010**. A rapid and inexpensive method for the direct PCR amplification of DNA from plants. *American Journal of Botany* **97**.

Bradshaw HD, Schemske DW. **2003**. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* **426**: 176–178.

Carlson JE, Holsinger KE. **2013**. Direct and indirect selection on floral pigmentation by pollinators and seed predators in a color polymorphic South African shrub. *Oecologia* **171**: 905–919.

Chase MW, Christenhusz MJM, Fay MF, *et al.* 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**: 1–20.

Coberly LC, Rausher MD. **2003**. Analysis of a chalcone synthase mutant in *Ipomoea purpurea* reveals a novel function for flavonoids: Amelioration of heat stress. *Molecular*

Ecology **12**: 1113–1124.

Crowden RK, Jarman SJ. **1976**. Anthocyanins in the genus *Erica*. *Phytochemistry* **15**: 1796–1797.

Dao TTH, Linthorst HJM, Verpoorte R. **2011**. Chalcone synthase and its functions in plant resistance. *Phytochemistry reviews : Proceedings of the Phytochemical Society of Europe* **10**: 397–412.

Gould KS, Dudle DA, Neufeld HS. **2010**. Why some stems are red: Cauline anthocyanins shield photosystem II against high light stress. *Journal of Experimental Botany* **61**: 2707–2717.

Gould KS, Kuhn DN, Lee DW, Oberbauer SF. **1995**. Why leaves are sometimes red. *Nature* **378**: 241–242.

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

Hao X, Horvath D, Chao W, Yang Y, Wang X, Xiao B. **2014**. Identification and evaluation of reliable reference genes for Quantitative Real-Time PCR analysis in tea plant (*Camellia sinensis* (L.) O. Kuntze). *International Journal of Molecular Sciences* **15**: 22155–22172.

Hoballah ME, Gübitz T, Stuurman J, *et al.* 2007. Single Gene-Mediated Shift in Pollinator Attraction in *Petunia*. *THE PLANT CELL ONLINE* **19**: 779–90.

Kearse M, Moir R, Wilson A, *et al.* 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28**: 1647–9.

Koes R, Verweij W, Quattrocchio F. **2005**. Flavonoids: a colorful model for the regulation

and evolution of biochemical pathways. *Trends in Plant Science* **10**: 236–42.

Landi M, Tattini M, Gould KS. **2015**. Multiple functional roles of anthocyanins in plant-environment interactions. *Environmental and Experimental Botany* **119**: 4–17.

Linder HP. **2003**. The radiation of the Cape flora, southern Africa. *Biological Review* **78**: 597–638.

Le Maitre NC. **2017**. Phylogenetics of the genus Erica and anthocyanin synthesis gene expression in Erica plukenetii. PhD Thesis. Stellenbosch University. http://scholar.sun.ac.za/handle/10019.1/101060.

Mugrabi De Kuppler AL, Fagúndez J, Bellstedt DU, Oliver EGH, Léon J, Pirie MD. **2015**. Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (Ericeae, Ericaceae). *Molecular Phylogenetics and Evolution* **88**: 121–131.

Nelson E, Oliver EGH. **2005**. Chromosome numbers in Erica - an updated checklist In: Nelson EC, ed. *Heathers 2*. Ipswitch: The Heather Society, 57–58.

van der Niet T, Pirie MD, Shuttleworth A, Johnson SD, Midgley JJ. **2014**. Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape shrub *Erica plukenetii*? *Annals of Botany*: 301–315.

Oliver EGH. **1989**. The Ericoideae and the southern African heathers. *Botanical Journal of the Linnean Society* **101**: 319–327.

Oliver EGH. **1999**. Systematic studies in the tribe Ericeae (Ericaceae–Ericoideae).

Pirie MD, Oliver EGH, Bellstedt DU. **2011**. A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggests numerous shifts in floral macro-morphology. *Molecular*

Phylogenetics and Evolution **61**: 593–601.

Pirie MD, Oliver EGH, Gehrke B, *et al.* 2017. Underestimated regional species diversity in the Cape Floristic Region revealed by phylogenetic analysis of the *Erica abietina/E. viscaria* clade (Ericaceae). *Botanical Journal of the Linnean Society* **184**.

Pirie MD, Oliver EGH, Mugrabi de Kuppler A, *et al.* 2016. The biodiversity hotspot as evolutionary hot-bed: spectacular radiation of *Erica* in the Cape Floristic Region. *BMC Evolutionary Biology* **16**: 190.

Platt AR, Woodhall RW, George AL. **2007**. Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques* **43**: 58–62.

Polashock J, Zelzion E, Fajardo D, *et al.* 2014. The American cranberry: first insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biology* **14**: 165.

Rausher MD. **2008**. Evolutionary transitions in floral color. *International Journal of Plant Sciences* **169**: 7–21.

Rozmer Z, Perjési P. **2016**. Naturally occurring chalcones and their biological activities. *Phytochemistry Reviews* **15**: 87–120.

Smillie RM, Hetherington SE. **1999**. Photoabatement by anthocyanin shields photosynthetic systems from light stress. *Photosynthetica* **36**: 451–463.

Streisfeld MA, Rausher MD. **2011**. Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution* **65**: 629–642.

Wessinger CA, Rausher MD. **2012**. Lessons from flower colour evolution on targets of selection. *Journal of Experimental Botany* **63**: 5741–5749.

Yuan Y-W, Sagawa JM, Frost L, Vela JP, Bradshaw HD. **2014**. Transcriptional control of

floral anthocyanin pigmentation in monkeyflowers (Mimulus). *New Phytologist* **204**: 1013–1027.

Zhang L, Xu P, Cai Y, *et al.* 2017. The draft genome assembly of *Rhododendron delavayi* Franch. var. *delavayi*. *GigaScience* **6**: 1–11.

Zhu Z, Wang H, Wang Y, *et al.* 2015. Characterization of the *cis* elements in the proximal promoter regions of the anthocyanin pathway genes reveals a common regulatory logic that governs pathway regulation. *Journal of Experimental Botany* **66**: 3775–3789.

Figure 1: *Chalcone synthase* (*Ep-CHS*), *chalcone isomerase* (*Ep-CHI*), *flavanone hydroxylase* (*Ep-F3H*), *flavanoid 3' hydroxylase* (*Ep-F3'H*), *dihydroflavanol-4-reductase* (*Ep-DFR*), *anthocyanin synthase* (*Ep-ANS*) and *UDP 3-O-glucosyltransferase* (*Ep-UDP-GST*) are the transcripts of the genes of the anthocyanin biosynthesis pathway enzymes required to produce red colour. The transcription of the enzymes' genes is collectively regulated by a complex of three transcription factors, *Ep-WDR*, *Ep-MYB* and *Ep-bHLH*. Expression of the genes of the anthocyanin biosynthesis pathway enzymes in eight growth points (1-8) of red-, pink-, and white flowered *E. plukenetii* ssp. *plukenetii* from the Du Toits Kloof Pass population as determined with Reverse Transcriptase PCR. *Ep-CHS* is not expressed in the white flowered individual. Expression of the other genes of the pathway and the transcription factors is otherwise essentially identical between red-, pink- and white flowered individuals. Samples were loaded on gels in the order of corolla length. A 1000 bp ladder was used.

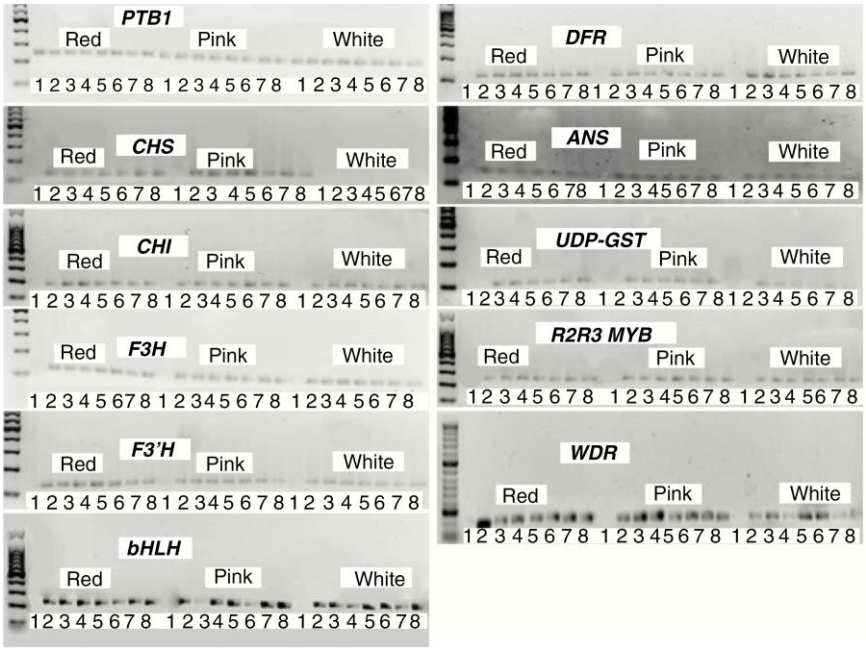


Figure 2: Expression of the genes of the anthocyanin biosynthesis pathway enzymes in red flowered *E. plukenetii* ssp. *plukenetii* from the Franschhoek Pass population as determined by Real Time Reverse Transcriptase quantitative PCR. Expression has been normalised to the expression of *Paratubulin 1*. Three biological replicates of each corolla growth point were tested.

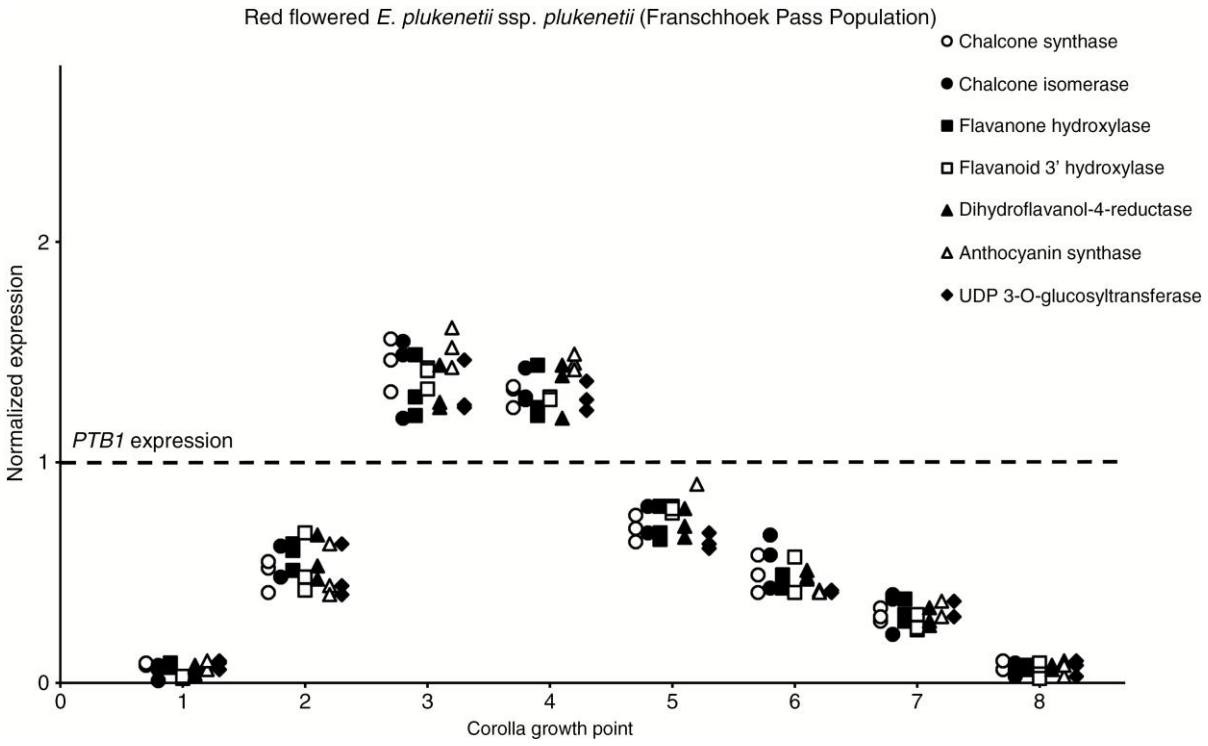


Figure 3: Expression of the genes of the anthocyanin biosynthesis pathway enzymes in white flowered *E. plukenetii* ssp. *plukenetii* from the Franschhoek Pass population as determined by Real Time Reverse Transcriptase quantitative PCR. Expression has been normalised to the expression of *Paratubulin 1*. Three biological replicates of each corolla growth point were tested. Expression of *Ep-ANS* is greatly reduced relative to the red flowered *E. plukenetii* ssp. *plukenetii* from the Franschhoek Pass population in Figure 2.

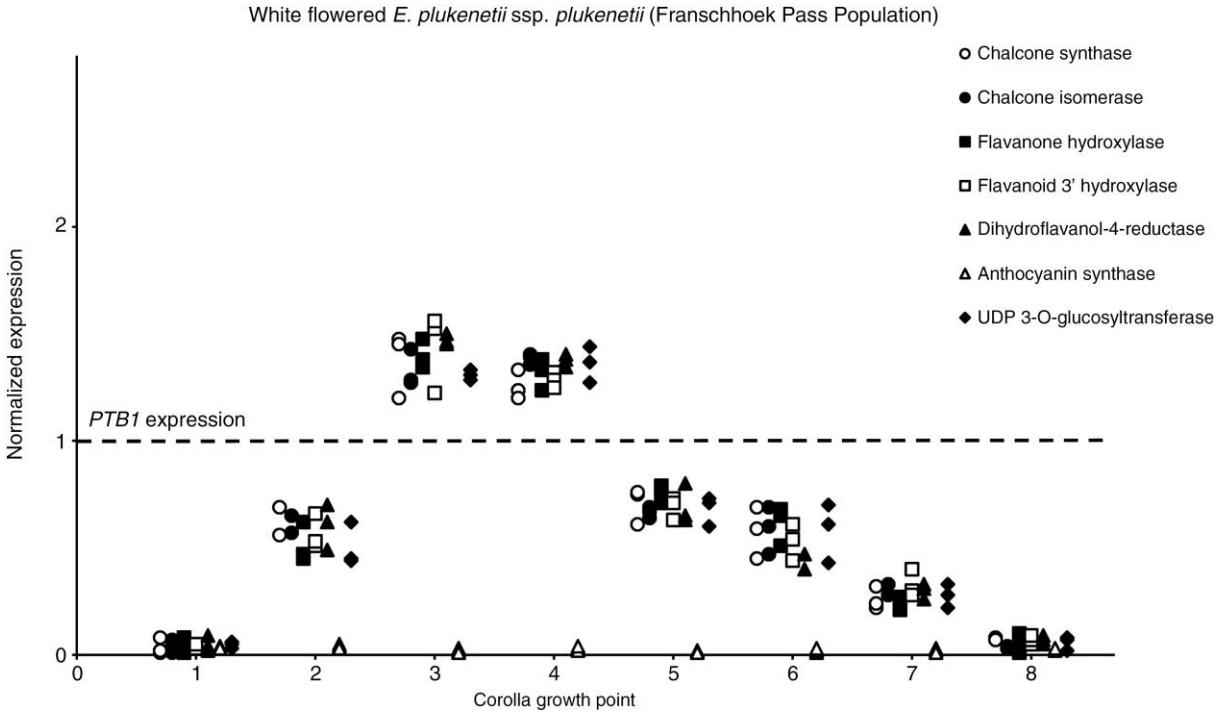


Figure 4: Expression of the genes of the anthocyanin biosynthesis pathway enzymes in white flowered *E. plukenetii* ssp. *breviflora* from the Witzenberg Valley population as determined by Real Time Reverse Transcriptase quantitative PCR. Expression has been normalised to the expression of *Paratubulin 1*. Three biological replicates of each corolla growth point were tested. Expression of *Ep-UDP-GST* and *Ep-DFR* is greatly reduced relative to the red flowered *E. plukenetii* ssp. *plukenetii* from the Franschhoek Pass population in Figure 2.

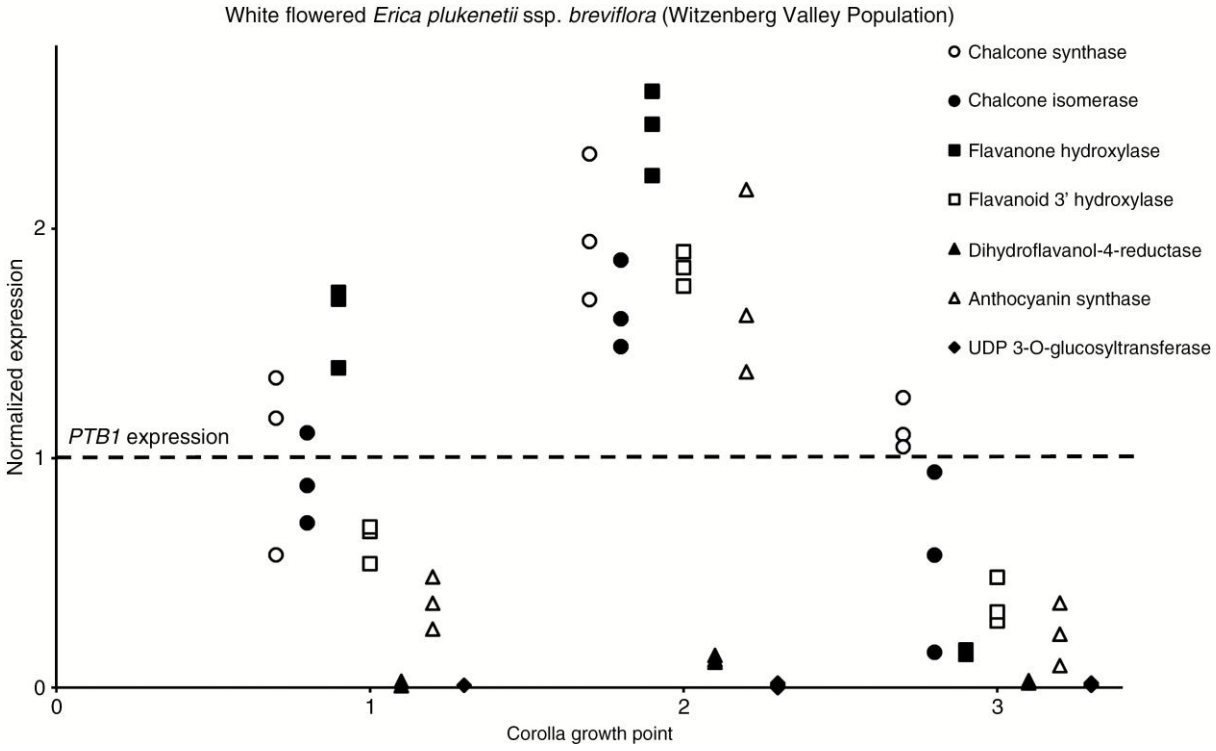
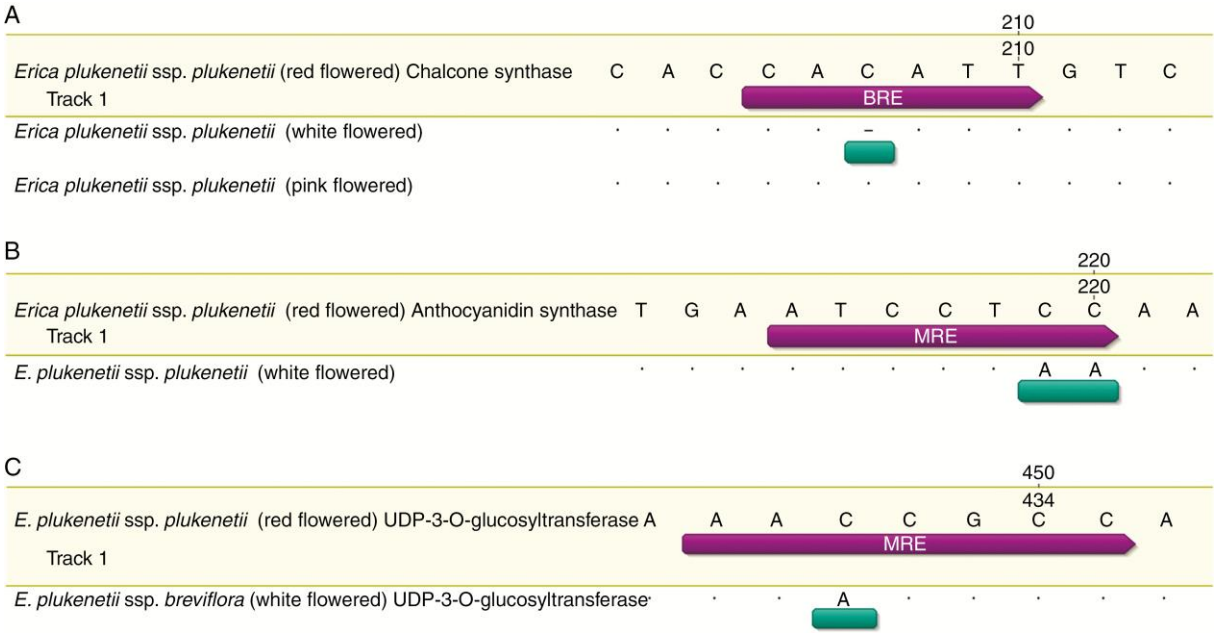


Figure 5: The mutations in the binding sites for the *trans*-acting regulatory elements that may account for the observed differences in gene expression between the red- and white flowered individuals. (A) The single base deletion in the bHLH recognition element of the *Ep-CHS* gene of the white flowered *E. plukenetii* ssp. *plukenetii* from the Du Toits Kloof Pass population. (B) The double mutation (CC to AA) in the MYB recognition element (MRE) of the *Ep-ANS* gene of the white flowered *E. plukenetii* ssp. *plukenetii* from the Franschhoek Pass population. (C) The single point mutation (C to A) in the MRE of the *Ep-UDP-GST* gene of white flowered *E. plukenetii* ssp. *breviflora* from the Witzenberg Valley population.



Supplementary Figure 1: Photographs of the different corolla growth points (1-8) of *Erica plukenetii* ssp. *plukenetii* at which anthocyanin sysnthesis enzyme gene expression levels were tested.

Supplementary Figure 2: (A) Red- and white flowered *Erica plukenetii* ssp. *plukenetii* flowering in the Franschhoek Pass population. (B) *Erica plukenetii* ssp. *breviflora* flowering in the Witzenberg Valley population.

65 Table 1: The primer sequences that were used to resequence the contiguous sequences from
66 the NGS data and also to amplify the region where transcription factors bind upstream of the
67 5’ UTR of the relevant gene of the anthocyanin biosynthesis pathway.

Name	Sequence	Annealing temperature (°C)
ANS-352F	GTCATTGACTTCCTCTTGCGC	
ANS-1,126R	TTGGGAATGGACTGGATGCC	60
ANS-1,023F	ATTTCGGGTACTGCCTGCAA	
ANS-2,269R	GAGTTCGTCCTTGCCACCA	66
ANS-2,175F	GAGGCCTAACGACCAAGGTC	
ANS-2,944R	AAGAGACAATCACAAGAGAAGTAGA	60
CHI-25F	AACCGACACAGCATCCAGAG	
CHI-878R	CGTGAAAGAAAGAACAAGAGGGT	60
CHI-632F	ACGGGCAAGCAATACTCAGA	
CHI-1,766R	CTTTGGGGAGGTCTTTGGCT	60
CHI-1,601F	ATTCCCAGTGAGCAACCACC	
CHI-2,522R	CAGGCTTTGAGTCCTCAGGG	60
CHI-2,411F	ACCACGGCAGCTTTATTCCA	
CHI-3,723R	GCCATGCTTCCCAACAATCG	60
CHI-3,646F	CCTGAGAAGGGCAAAGTGGT	
CHI-4,546R	CAAAGCTTCCACAGTATGCCA	60
CHS-9F	ATACGTTCCTGGCTACCCCT	
CHS-1,097R	ACTCCTTGATGGCCTTCACG	60
CHS-992F	AATGTGTGCGCATACATGGC	
CHS-2,020R	ACTGCAAACAACGGGCCTAA	60
DFR-2,142F	GTGGGAGTAGAGTAGCCCCA	
DFR-2,866R	TTGGTTGAGGGCACAAACCT	62
DFR-2,720F	AGGATAACGTGAACGGCTCG	
DFR-3,500R	CCGGACGGAGGGAGTAGTTA	66
DFR-3,351F	AGCTTTGACGAGGCCATTGA	66

DFR-4,242R	GGTTCAAGTGCTGCCCTACT	
DFR-4,722F	ACGATCTCAAACTCAGGGCC	
DFR-5,823R	TGCTCTTGGACATTGACGGT	56
DFR-5,714F	AAGCCGACGATCAATGGTGT	
DFR-6,920R	CCCCATCCCTTGCAACTTCT	60
DFR-6,044F	GAGCATCCTGAAGCAGAGGG	
DFR-6,996R	TTGTTTGAGACTGCTTTATATTTTCG	62
F3'5'H-261F	CGGAGATGCTCACGTACTCC	
F3'5'H-1,158R	AGCTCGTGCAAACTGACCAT	60
F3'5'H-1,077F	GGTTGTGGAAAGTGAACGCA	
F3'5'H-2,274R	ACCACCACATTGACCATCGA	60
F3'H-1,335F	TCGGGCAAATATAACCTCGGA	
F3'H-2,134R	AAGTCGTCCAAGGCCTTAGC	60
F3'H-2,009F	CTCCGGGGCCAAGCATATT	
F3'H-3,130R	ACAATAATGCTAGCCCCGGG	60
F3'H-3,050F	TGGAAAACGAACCATGCGTTC	
F3'H-4,027R	ACCGATGAGTTAAGTTGGCGA	60
F3'H-3,926F	TGGTGAACAGACTTTTCAACTGG	
F3'H-5,083R	GGTGCACAGGTTAAGGAGCT	60
F3'H-5,000F	GAAGTAGCCATTCTGACGCG	
F3'H-6,088R	CCCATTACGGTTGAGCACGA	60
F3'H-6,070F	CGTGCTCAACCGTAATGGGA	
F3'H-6,487R	AGGAGCGCATGAAACCGTAT	60
F3'H-6,287F	CGATCTGAGACCCAACCCAA	
F3'H-7,426R	GGAGAGAGAGTGGGGTGGAT	60
F3'H-7,230F	CTCGTCCAGCACAGTGGAAT	
F3'H-8,391R	AAGTGAACTTTTCATCCCCTTTTT	60

F3H-528F	CGTTTAGACCTTCTCTCGAGCA	
F3H-1,519R	ACAATCTTAAGTTTCCCATATTGACCT	60
F3H-1,060F	CAGTCAAAGTTCGTCCGGGA	
F3H-2,139R	GCGATGCGATCCGGTTTAAC	60
F3H-2,023F	AAGCGGTCCCACAGTTTTGA	
F3H-3,740R	ATGACCATGGTCGCCCAAAT	56
F3H-2,810F	GAGCGAGAGTAAGTTGCCGT	
F3H-3,740R	ATGACCATGGTCGCCCAAAT	50
F3H-3,562F	AATGTCCACAACCCGACCTC	
F3H-4,617R	ACGTACGGGATTGGTGCTTT	60
F3H-4,448F	GCCCAGGACAGCCCAATTAT	
F3H-5,794R	TGGTGGGTAGCAAATCTCGG	60
UDP-GST-38F	TCGGGGTAGGTTTTCGTGTG	
UDP-GST-839R	GTTGAACTTTGCGGCGACTT	60
UDP-GST-720F	TTGAAGGCCACCATGATGCT	
UDP-GST-1,884R	CGTTGTCATCTCACGCCTGA	60
MYB-26F	GCGTCCACTTGTGTGTTTCC	60
MYB-1,395R	AACACGTACGGCTAACACAA	
bHLH-353F	GACCCATTGACGTGTTTGCC	60
bHLH-1,360R	ACAGACAACAGAGCTTCACA	
WDR-707F	CAGGTTGCCTTGTTTGCAGT	60
WDR-2,220R	TGTGATGAACAATGTGGGGC	

68

69

Table 2: The primer and probe sequences that were used for RT-PCR and RT-qPCR that were used to determine the expression of the genes of the anthocyanin biosynthesis pathway in *Erica plukenetii*.

Name	Label	Quencher	Sequence	Expected Product size (bp)
ANS-1,281F			AGTCCTCTCCCTAGGCTTGG	
ANS-1,328Probe	6-FAM	3' Iowa Black® FQ	AAGTTGGTGGCAAGGACGAA	167
ANS-1,448R			ATGAAGGTGAGGGCGCTTAC	
CHI-632F			ACGGGCAAGCAATACTCAGA	
CHI-657Probe	Hex	3' Iowa Black® FQ	TGGTGGAAAACTGTGTTGCC	184
CHI-816R			CTAACCGTTAGCGACCCCAG	
CHS-153F			CCGTCATGGCTATCGGGAC	
CHS-188Probe	Cy5	3' Iowa Black® RQ-Sp	TGCGTTGATCAGGCCACTTA	109
CHS-262R			CTCCTTCAACTCGGCCTTGT	
DFR-112F			AGGATAACGTGAACGGCTCG	
DFR-175Probe	Cy5	3' Iowa Black® RQ-Sp	GCTCCTGGCTCATCATGAGG	117
DFR-229R			ACGGTGGCTCGAACAACATA	
F3'5'H-261F			CGGAGATGCTCACGTACTCC	
F3'5'H-360Probe	Hex	3' Iowa Black® FQ	ACATGGTGGTGGAGCTCATG	139
F3'5'H-400R			GTTGAATAAACCGGCCGACG	
F3'H-406F			CTCCGGGGCCAAGCATATT	
F3'H-471Probe	Cy5	3' Iowa Black® RQ-Sp	GGCGGATGCTCAGGAAGATA	125
F3'H-531R			AAGTCGTCCAAGGCCTTAGC	
F3H-259F			GATATCGCTAGCCGGGATCG	
F3H-338Probe	6-FAM	3' Iowa Black® FQ	TGGGGGATATTCCAGGTGGT	127
F3H-386R			TAATCAGACCGGCATCCACG	

UDP-GST-480F			AAGTCGCCGCAAAGTTCAAC	
UDP-GST-512Probe	Cy5	3' Iowa Black® RQ-Sp	GTTTTCCACGGCATCAGCTT	135
UDP-GST-615R			GGCACCAAAAAGGGTTCGTC	
PTB1-1,416F			TTCATCAGAACCGGCTCAGG	
PTB1-1,500Probe	Texas Red	3' Iowa Black® RQ-Sp	CGCTGATGTCGCTGGAAATG	148
PTB1-1,564R			TGCTGACAAGACGTGCATCA	
MYB-998F			ATAACCCAAAGCCCACGAGG	136
MYB-1,134R			CACCCGATCAACCTCAGCTT	
bHLH-645F			AGTTGCGGAGGGATAGGCTA	151
bHLH-796R			GTCTGTTCTGGGAGGCCTTC	
WDR-1,605F			CAGGACCCCAGGTATACGGA	151
WDR-1,756R			CCTCACTCGCACTGTGGAAT	

73