Transcriptomics: Open Access

Research Article **Open Access**

Characterization and Expression Analyses of Chalcone Synthase (CHS) and Anthocyanidin Synthase (ANS) Genes in *Clivia miniata*

Achilonu C Conrad* and Maleka F Mathabatha

Department of Genetics, University of the Free State, Republic of South Africa

Abstract

Chalcone synthase (CHS) and Anthocyanidin synthase (ANS) are amongst the key enzymes responsible for the production of anthocyanins in plants. They are generally encoded by multi-gene families with some members of these families contributing to colour pigmentation. The study examined CHS and ANS genes in Clivia miniata; whose flowering tissues undergo colour changes. The RNA to cDNA was isolated from floral tissues and three unigene(s) (*CmiCHS 11996*, *CmiCHS 43839* and *CmiANS*) of short sequence lengths were initially obtained using next-generation sequencing technique. Gene-specific primers were designed from the unpublished initial short sequences of CHS and ANS. The length of the amplified cDNA after PCR were *CmiCHS 11996* (933 bp), *CmiCHS 43839* (951 bp) and *CmiANS* (983 bp). The start translated ORF frame corresponding to the predicted protein of 390 amino acid deduced protein (AEN04070) for CHS genes, and predicted 355 amino acid in respect to ANS gene (AGD99672). *In silico* analysis revealed the calculated molecular weight and theoretical isoelectric point (pI) *CmiCHS 11996* and *CmiCHS43839* were 31.0 kDa - 6.95 and 34.6 kDa - 7.54 respectively. The important motifs of the product binding site and active site were successfully identified from the deduced amino acid sequences. Multiple sequence alignment showed that the *CmiCHS* and *CmiANS* sequences were highly conserved and shared high sequence identity (>83%) with chalcone synthases from other plants. However, another assay was performed to determine the expression profiles of these genes in different tissues as well as the tepals (orange and yellow flower) using the real-time quantitative PCR. The expression levels of *CmiCHS* and *CmiANS* were higher in tepals compared to other tissues (leaves, style and stigma and scape). These expression patterns of the genes in the tissues corresponded to the accumulation of anthocyanin, suggesting that the orange and yellow colour pigments was certainly related to the regulation of chalcone synthase and anthocyanin synthase.

Introduction

Amaryllidaceae (family name) originated from a flowering plant which genus Clivia falls under the category. However, Clivia are herbaceous evergreen with a strap-like leaves flowering monocot plant that is endemic to Southern Africa [1]. The morphology of each of the flower comprises of three sepal and petals, collectively known as tepal - which is attached to the base of the plant. Clivia flower mainly contributes to a range of colours from yellow through orange to red. However, the plants consist of six species, namely *C. nobilis* [2], *C. miniata* [3], *C. gardenia* [4], *C. caulescens* [5], *C. Mirabilis* [6] and *C. robusta* [7]. Amongst these species, *Clivia miniata* Regel (*C. miniata*) [8] are the most commonly cultivated species in various parts of the world, mostly in Australia, China, Belgium, Japan, New Zealand and USA.

The values of so many varieties of flower colour are determined by a major factor in flower pigmentation during flavonoid biosynthesis. The pigmentation of flowers which serves as an ornamental plant value are due to the composition of plant secondary metabolites known as betalains, carotenoids and flavonoids [9,10]. These individual metabolites triggers different functions to plants; betalains are nitrogen-containing compounds which originated from tyrosine, these produces violet to red and yellow to orange colouration to flowers and are uniquely found in Caryophyllales [11,12]. Carotenoids are responsible for flower pigmentation with different colour ranging from yellow to red and also vital for photosynthesis. While flavonoids are important and most widely spread pigments in the plant taxon and they hugely involved in diverse pigmentation from orange to pink, red, violet and blue colours in fruits and flowers [13,14].

In addition, flavonoids such as chalcone, flavones, flavonols and anthocyanins have varieties of biological functions which include, the protection of cells against UV radiation, molecule signaling in plants and microbes interaction, resistance against pathogens and herbivores, root nodule organogenesis, and pollen growth & development respectively [15-18]. The accumulation of plant secondary metabolites influence flower pigmentation and these include carotenoids, flavonoids and betalains [15]. Flavonoids have different classes that are activated in many plant functions. These functions are the protection against UV light (flavonols), defence against pathogens (isoflavonoids), and plant pigmentation (anthocyanins) [19]. There are enzymes such as chalcone synthase and anthocyanidin synthase are involved in catalyzing reactions during anthocyanin biosynthesis which shows different characteristics in substrate specificity in plants [20].

Chalcone synthase are enzyme initiator in the anthocyanin biosynthesis and it catalyzes the condensation of a molecule such as phenylpropanoid CoA-ester (e.g. P-coumaroyl-CoA) with three C_2 units from malonyl-CoA, and cyclizes the resulting tetraketide intermediate to yield a chalcone (e.g. Naringenin chalcone) [21]. In addition, Grotewold [22] stated chalcone are isomerized quickly by the chalcone isomerase (CHI), which leads to the synthesis of naringenin. Naringenin is later converted to other enzyme branches that lead to the production of anthocyanins. Furthermore, CHS contains a multi-gene family, which offers a vital model to comprehend the functional significance and various patterns of expression levels of the gene family members. This gene is structurally conserved across taxa and most of

them consist of two exons separated by one intron. Antirrhinum is an exceptional case with two introns between the two exons [23,24]. Chalcone synthase gene families such as Vitis, Petunia, Ipomoea, Arabidopsis, Ginkgo, and Tulip have also been isolated from various plants [12,23,25-28].

However, anthocyanidin synthase are also key enzymes in the synthesis of anthocyanins, follows dihydroflavonol reductase (DRF) pathway and the conversion of colourless leucoanthocyanidins to coloured anthocyanidins [29]. The proteins of ANS is derived from 2-oxoglutarate-dependent dioxygenases (2-ODDs) family proteins that incorporates the flavonoid enzymes such as flavanone 3-β hydroxylase (FSH), flavonol synthase (FLS) and flavone synthase I (FS I) proteins. These proteins use molecular oxygen as the co-substrate and they are characterized by their co-factors [30]. Research of ANS structural gene affecting the production of anthocyanin in other plants have being done. In onion (*Allium cepa*), the genes are classified into two groups. The first category represents regulatory genes which control the expression of the structural genes, and this regulatory process can be affected by internal and external physiological stimuli. The second category of the structural genes encodes enzymes in the anthocyanin biosynthetic pathway which contributes to the colour pigmentation [31]. However, the characterization of the structural genes of ANS isolated in *Allium cepa* contains six exons and five introns. It differs from other plants such as in *Ginkgo biloba* (*GbANS*) gene with three exons and two introns [32] and in tomato (*Solanum lycopersicum*)' three exons and two introns unpublished (Bombarely and Mueller).

Understanding the molecular background of Clivia is vital; and currently there is paucity of information regarding flavonoid biosynthesis of this genus. There have been few studies of CHS and ANS genes done in Clivia species but there're no reports so far regarding the full-length characterization of these genes in Clivia species. In this paper, we describe the use of RACE to amplify full-length cDNA with a gene-specific-primer (GSP). The characterization of *cmiCHS* and *cmiANS* putative genes and their amino acids sequences were compared with other plants using the required analytical tools. These results give insight into molecular approaches used for characterizing the CHS and ANS gene family in Clivia. Notwithstanding to also ascertain the pigmentation and regulatory mechanism of chalcone synthesis in Clivia flower, we compared the expression profiles of three putative chalcone synthase *CHS* genes and one anthocyanidin synthase gene (*CmiCHS 11996, CmiCHS 43839, CmiCHS 50130* and *CmiANS 59543*) in different *C. miniata* cultivar tissues (*C. miniata var. miniata and C. miniata* var. citrina) from roots, leaves, scape's, style, stigma, berry and the tepals-at two different developmental stages. This study gives insight into understanding the molecular approaches in colour manipulations and exploring and analyzing single nucleotide polymorphism (SNP) in Clivia. These could pave way into translational biology via developing biotechnological strategies for enhancing mutations and gene modification in *Clivia* genome in the future.

Materials and Methods

Plant materials

The Clivia cultivar, *C. miniata* var. *miniata* '*Teleurstelling*' (orange flower) and C. *miniata* var. citrina 'Giddy' (yellow flower) was grown under this condition 13°C to 27°C days and nights in the Glass House of the Department of Genetics, University of the Free State, Bloemfontein, South Africa. The orange flower tissue tepals were sampled at the final developmental stages in summer (full pigmentation). In addition, these sampling was done in a day and used for a three times repetitive characterization assay. While two biological groups of leaves stem, style,

stigma, and berries tissues were collected and processed. The tepals of both orange and yellow flowers were collected at two different flower developmental stages defined as follows: stage 1-un-pigmented and stage 5-open matured flower fully pigmented and stage. Furthermore, the assay on gene expression for two biological groups was done in triplicates.

Total RNA isolation and cDNA synthesis

Fresh tepals samples (≈0.5 g) of each were homogenized in liquid nitrogen to a fine powder using a sterile pro-cooled mortar and pestle. An approximate biomass of 0.1 g of each sample was suspended in Trizol˙ (Invitrogen). After the addition of Trizol˙, the suspension was incubated for 5°C min at 30°C using the Heating Block (AccuBLOCK™ Digital Dry Bath, Labnet Int'l, Inc). Afterwards, chloroform was added to the sample, the mixture was vortex for 15 sec, and incubated at room temperature for 3 min. The samples were then centrifuged at no more than 12,000 g for 20 min at 4°C. Following centrifugation, the sample mixture was separated into phenol-chloroform (organic) phase, an interphase and the upper aqueous phase, which contained the RNA (RNA Ribonucleic Acid).

The supernatant was discarded and the RNA was precipitated by adding isopropanol, followed by vortex and incubation for 10 min at room temperature. The precipitation was centrifuged at 12, 000 g for 15 min at 4°C, the supernatant discarded and the precipitate washed with 500 µl of 75% ethanol. The RNA pellet was briefly (air)-dried, and re-suspended in 50 µl DEPC-treated (RNase-free) water and then incubated at 60°C for 10 min. The concentration of the RNA samples was measured using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). At this point only the RNA samples with a 260/280 nm absorbance ratio (shows an evidence of protein contamination) between 1.9 and 2.1, were used for first strand cDNA synthesis. The quality of the RNA samples was determined by electrophoresis on an Agarose gel of 2% with ethidium bromide staining. The RNA extract was kept at -80°C until further use.

Reverse transcriptase polymerase chain reaction (RT-PCR) was used to synthesize first-strand cDNA from RNA. In preparation of the reaction, 1 µg RNA and (100 µM) anchored Oligo (dT)$_{18}$, dNTP (Deoxynucleotide Triphosphate) mix (10 mM each) (Thermo-Scientific) was added, followed by addition of DEPC-treated water to a final volume. The solution was mixed gently, pulse centrifuged, incubated at 70°C for 5 min and chilled on ice. Afterwards, the following reagents were added: 5 × RT buffer (250 mM Tris-HCl, pH 8.3; 250 mM KCl; 20 mM MgCl$_2$; 50 mM DTT), 20 U of RiboLock® RNase inhibitor was added (Thermo Scientific) and DEPC-treated water were added to the required volume. The reaction mixture was incubated at 37°C for 5 min after which a required volume of RevertAid™ M-MuLV Reverse Transcriptase (200 U) (Thermo scientific) was added. The incubation step at 42°C for 60 sec was performed followed by heating at 70°C for 10 min to inactivate the reaction and then cooled on ice. The heating steps were performed in a thermal cycler 2720 (Applied Biosystems). The cDNA products were diluted to a ratio of 1:50 (cDNA: dH$_2$O).

Primer design and normalization of reference genes

Two gene-specific primer sets were designed using Primer Designer 4 Software and the other 2 sets of CHS designated primers were designed by aligning two CHS sequences of a monocot species highly identical to *Clivia miniata* assigned in the GenBank. The DNA sequence of *Narcissus tazetta* var. chinensis cultivar white (JN227883) were used to design primers via online web page of Integrated DNA Technologies 'www.eu.idtdna.com'. A primer map was generated and the appropriate CHS primer sets were selected (CHS 3 and CHS 5). While ANS gene-specific

primers were designed from an obtained partial sequence Maleka et al. (unpublished) using online tool of the Integrated DNA Technologies. The 'Oligonucleotide Properties Calculator' online tool was used to determine the appropriate melting temperature (Tm), GC% content and any possible formation of Hairpin's for all primers (Table 1). The primers were designed according the following standards: a Tm closest to 60°C, a GC content of at least 45%, a primer length of not less than 20 nucleotides and a minimum PCR product size of 700 bp.

A total of six reference genes were selected based on their stability in Clivia tissues and other plants; searched and available in the database "NCBI" (National Centre of Biotechnology and Information technology). These reference genes are homologous sequences obtained from approximately 20 ESTs and were used for designing primers for the amplification of the six reference genes, while the four target genes from partial putative sequences obtained were chosen for this study Maleka et al. "unpublished" (Tables 2 and 3). Computer simulation was performed to design gene-specific primers (GSPs) for real time qPCR using IDT® tool "Integrated DNA Technologies". The primers were designed from partial putative gene sequences of Chalcone synthase and Anthocyanidin synthase obtained from the analyses of *C. miniata*. The primer parameters were taken into consideration by performing an 'Oligonucleotide Properties Calculator'–a website tool was used to determine the appropriate melting temperature (Tm), GC% content and any possible formation of Hairpin's for all primers. The primers were designed according the following standards: a Tm closest to 60°C, a GC content of at least 45%, a primer length of not less than 19 nucleotides and a minimum PCR product size of 85 bp

The primer sets of both the reference and target genes were used to amplify fragments in *C. miniata*, which were designed to have certain characteristics. All primers had a preferable length of between 19 to 22 nucleotide bases. The temperature (Tm) of each primer in a primer pair were very similar, which differs with 1°C or 2°C amongst them. All primer sequences were submitted to "IDT" for commercial synthesis. Once the primers were delivered in a pellet form in a tube, concentrated stock solutions of 100 μM (100 pmol/ μl) were prepared by dissolving each lyophilized pellet in Tris EDTA (TE) buffer (10 mMTris, pH 7.5; 1 mM EDTA, pH 8.0). 0.5 μM working solutions were prepared for future use during teal-time quantification PCR.

To normalize the reference genes, the amplification reaction consisted of 2 units SYBR Green I Master Mix (Applied Biosystem), 0.1 μM of both primers and 2 μl of cDNA tissues (in triplicates). The cycling conditions were 2 min at 50°C, 10 min at 95°C and 40 cycles of 15 sec at 95°C and 1 min at 60°C. The melt curve analysis cycling conditions were 15 sec at 95°C, 1 min at 60°C followed by ramping from 60°C to 95°C. The Ct- values were averaged and were imported to Normfinder Excel Sheet tool; used for reference gene selection.

PCR amplification of the unigene(s) and candidate reference genes

The reaction was done using gradient PCR program to determine its perfect annealing temperature. The PCR reactions with the resulting cDNA template were carried out in 6 PCR tubes of forward and reverse primer (10 μM) added, 2 ng of first-strand cDNA, HiFiReadyMix DNA

Gene	Primer	Sequence (5' - 3')	Primer Length (nt's)	Tm (°C)	Tm_a(°C)	Ampli. Length (bp)
GSP11996	F	GCATCGAGCAGAGTGCTTATC	21	61	58.5	963
CHS3	R	GATCCAGAACAACGAGTTCCA	21	60		
GSP43839	F	TATCGAGCAGAGTGCTTATCC	21	61	58.5	1082
CHS5	R	GCCCTCATCTTCTCTTCCTTG	21	62		
cmiANS	F	GTGCTAGAGGAGGTGAAGAAAG	22	62	60	720
	R	GAGACCCACTTGTCGTTGTAG	21	62		

(bp)=base pair, (Tm/°C)=Temperature in degree Celsius, (Tm$_a$)=Annealing temperature, GSP=Gene-specific primer, (Ampli. length)=Amplicon length, (F)=Forward, (R)=Reverse.

Table 1: Gene-specific primers for gene amplification.

Symbol	Primer (5' - 3')	Primer length (bp)	Primer Tm (°C)	Ta (°C)	Amplicon length (bp)
18s	F:CGCGCTACACTGATGTATTC	20	60.4	61	130
	R:CTGATGACTCGCGCTTACTA	20	60.4		
EF1α	F:TGATTGCCACACTTCTCACATTG	23	60.99	61	148
	R:ACCATACCAGCATCACCATTCT	22	60.81		
PP2A	F:CAGGTGCTGGACAAGGTCAA	20	62.45	61	226
	R:TCGTCTCCTCCACCACAGAA	20	62.45		

(18s) 18s ribosomal RNA,(EF1α) Elongation factor 1- alpha,(PP2A) Protein phosphate-2A, (αTUB) alpha-Tubulin, (G6PDH) Glucose-6-phosphate dehydrogenase, (RNP2) small nuclear ribonucleoprotein-2, (UBQ) Ubiquitin, (bp)=base pair, (Tm/°C)=Temperature in degree Celsius, (Ta)=Annealing temperature, (F)=Forward, (R)=Reverse, (n/a)=Not available, (Eff.%)= Efficiency percentage, (R²)=Regression coefficient.

Table 2: Reference gene primer sequences used for gene expression assay.

Symbol	Primer (5' - 3')	Primer length (bp)	Primer Tm (°C)	Ta (°C)	Amplicon length (bp)
cm11996	F:GAGTATCGATGAGATTCGCAAGA	23	62	57	98
	R:CACTCTGCTCGATGCAATTTAG	22	61		
cm43839	F:GACCGGCTACTATCTTAGCTATTG	24	62	57	102
	R:GTGTTCGCTATTTGTGACACG	22	62		
cm50130	F:GCATGTGCGAGAAGTCTATGA	21	62	57	120
	R:GGAGACATGTAGGCGCAAA	19	62		
cmiANS	F:GGCTCGTGAACAAGGAGAAGGT	22	64.54	61	105
	R:GCGGCCTGAGGACGATCTTA	20	64.5		

(bp)=base pair, (Tm/°C)=Temperature in degree Celsius, (T$_a$)=Annealing temperature, (F)=Forward, (R)=Reverse, (Eff. %)=Efficiency percentage, (R²)=Regression coefficient.

Table 3: Target gene primer sequences used for gene expression assay.

Polymerase (2 U/μl) (KAPA Biosystems) and sterile 1x distilled water (dH$_2$O) to a final volume. Gradient PCR was performed in a thermal cycler "G-Storm Labtech GS04822" (Labtech International Ltd) using the GSP with the CHS primers sets; and the amplification program was carried out with an initial denaturation at 95°C for 3 min, followed by 35 cycles of 98°C for 20 sec, (Tm1 49.8°C, Tm2 50.2°C, Tm3 51.4°C, Tm4 53.5°C, Tm5 56.4°C, Tm6 58.5°C, Tm7 59.6°C, Tm8 59.7°C), range of 10°C for 30 sec, 72°C for 120 sec. A final extension step at 72°C for 4 min followed; after which each PCR mixture was stored at 4°C (for immediate use) or at -20°C (for long-term use). Furthermore, the above PCR program was applied using a specific annealing temperature of 58.5°C and 60°C of the successful amplification (Table 1) were chosen for further analysis after the gradient PCR products were viewed on an Agarose gel.

Gradient PCR analysis were carried out in a QuantStudio-5 RT-qPCR instrument (Applied Biosystems). The reference genes and target genes were amplified using the cDNA tissues (Orange flower stage 5) was performed. Reaction mixture consisted of (2 μg template cDNA, 2X SYBR Green I Master Mix (Applied Biosystems), 0.1 μM of both primers. The cycling conditions of the PCR products were (2 min 50°C, 10 min 95°C and 40 cycles of 15 sec 95°C and 1 min 60°C). Then followed by the melting curve cycling conditions (15 sec 95°C, 1 min 60°C, a ramping step from 60°C to 95°C, a ramp speed of 2% and the final step of 15 sec at 95°C. The gradient PCR products were ran on a 2% Agarose gels at 100 Volts for 120 min using electrophoresis and assessed by visualizing it on a 'VacutecSyngene G-box Vacutech. The specific annealing temperature was consistently used onwards during the experiment to determine the standard of the reference genes and target genes.

Quantitative real-time PCR (RT-qPCR)

In the interest of determining the expression pattern of Chalcone synthase gene in *C. miniata*, the quantitative RT-qPCR was carried out in a 96-well plate using an ABI 7500 Real-Time PCR SDS (Sequence Detection System)–Applied Biosystem. The real-time qPCR was performed using SYBR® Green I detection chemistry, and the assay was performed according to the manufacturer's instructions; done in the required final volume which contains the SYBR® FAST qPCR Kit-Master Mix (2X) ABI PrismTM (KAPA Biosystem), 0.2 μM of each specific primer, 2 μg template cDNA and nuclease-free water. The primer sets without the cDNA were prepared on each reaction plate in triplicates and designated as 'no template control' (NTC). In order to ascertain the efficiency of the RT-qPCR amplification performed on the orange flowers, reactions for a standard dilution series was performed on the cDNA samples (template) for each *CmiCHS* and *CmiANS* gene(s) targeting the reference genes.

The RT-qPCR was carried out according to the following thermal profile cycling conditions (enzyme activation- 3 min at 95°C, followed by 45 cycles of denaturation for 15 sec at 95°C, annealing/extension for 40 sec at 60°C. In addition, the melting curve was generated at the end of the cycles used to evaluate the specificity of the amplification; while the baseline and threshold cycles (Ct) were determined using the QuantStudio-5 (Applied Biosystems). However, primer efficiencies and standard deviations were calculated based on a standard curve generated. The quantification of each gene expression for CHS and ANS at all different tissues were calculated.

Sequence analysis of *CmiCHS* and *CmiANS* PCR products

The gradient PCR samples of the cDNA fragments with fine bands were cut off and purified with a purification kit (BIoFlux gel extraction kit; Bioer). A cycling sequencing reaction volume was prepared for each template with a BigDye terminator v3.1 kit (Applied Biosystems). The reaction component of Terminator Premix, sequencing primer (3.2 μM), dilution buffer, and 2 ng template DNA (which had a final concentration of ≈10 ng/μl in the sequencing mixture) were added. The reactions were performed in a thermal cycler "G-Storm Labtech GS04822" under the following conditions: initial denaturation at 96°C for 1 min, followed by 25 cycles of 96°C for 10 sec, 50°C for 5 sec, 60°C for 4 min, and then storage at 4°C.

The post-reaction clean-up was done with EDTA/Ethanol precipitation. Each sequencing reaction volume was transferred to a 1.5 ml Eppendorf tube; it adjusted to a final volume with 1x dH$_2$O, 125 mM EDTA and absolute ethanol, followed by 5 sec vortex and then precipitation at room temperature for 15 min. The samples were centrifuged at 20 000 g for 15 min at 4°C and the supernatant was completely aspirated. The pellets were washed with 70% ethanol by vortex briefly and centrifugation at 20,000 g for 8 min at 4°C. The supernatant was completely aspirated and the pellet was air-dried for 20 min. The samples were then stored in the dark at 4°C until analysis using an "Applied Biosystems 3130 Genetic Analyser" (Applied Biosystem).

Data Analysis

The nucleotide sequence analysis and the multiple sequence alignment were performed by Clustal Omega EMBL-EBI online tool using default parameters (www.ebi.ac.uk/Tools/msa/clustalo/), while the protein domains and active sites were determined using InterPro-EMBL-EBI Server (www.ebi.ac.uk/interpro/), while Signal P 3.0 Server (www.cbs.dtu.dk/services/SignalP/) was used to predict signal peptide from the deduced amino acid sequences. A pair-wise alignment was done on the obtained nucleotide sequences using LALIGN - Readseq version 2.1.30 to deduce any multiple matching sub-segments in the two sequences (embnet.vital-it.ch/software/LALIGN_form.html). ProtParam tool was used to determine the physical and chemical parameters of the proteins such as; the theoretical isoelectric point (pI) and molecular weight (Mw) (web.expasy.org/protparam/), and the prediction of transmembrane helices in proteins was performed by TMHMM Server v. 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/). Protein classification was performed with Superfamily 1.75 (supfam.cs.bris.ac.uk/SUPERFAMILY/). The three-dimensional (3D) protein structure modeling homology of the *CmiCHS* was performed using YASARA-MODEL software tool (www.yasara.org/), and phylogenetic trees were constructed by MEGA software using multiple aligned sequences of CHS and ANS plants.

The Ct values (cycle threshold) derived from the RT-qPCR results were compiled using the comparative ΔΔCt method [33,34]. The experiment was performed in triplicate for all the 14 cDNA tissues of the two biological plant groups and the data showed the Ct mean values. The raw Ct data were exported to a spread sheet and the $2^{-\Delta\Delta Ct}$ for the relative quantitative expression values for both target and reference genes were generated. It was obtained by subtracting the calibrator ΔCt value from the ΔCt of the samples [33,35]. Samples of highest ΔCt value or lowest expression level were used as the calibrator. Further analysis of determining the log10 of the $2^{-\Delta\Delta Ct}$ values; the log10 values were plotted against the different corresponding tissues using Microsoft Excel® 2013.

Phylogenetic analysis

Multiple sequence alignment of the amino acid sequences from higher plants of CHS and ANS superfamily protein were separately

compared and aligned with ClustalO program; using the default parameter settings (www.ebi.ac.uk/Tools/msa/clustalo). Further analysis was done to construct the phylogenetic tree, the aligned sequences was saved and carried out by MEGA software tool. It generated a neighbour-joining tree with bootstrapping (1000 replicates) analysis; gaps were considered with pairwise deletion.

Results

Sampling of tissues

In Clivia flower, there are known 5 developmental stages in general. The unpigmented bud (stage 1), slight pigmentation appears (stage 2), pigmentation appears over one third of the surface (stage 3), pigmentation covers about two third of the surface with buds before the anthesis (stage 4) and visible and matured-fully pigmented flower (stage 5). This study required only stage 1 and stage 5 of both orange and yellow flower with other different flowering tissues 'leaves, stem, style and stigma, berries' (Supplementary information 1)-in order to see a differential expression pattern in the flowering tissues compared to the rest of the plant tissues [36,37]. Therefore, this study quantified the expression of *CHS* (chalcone synthase) genes been an early stage in flavonoid biosynthetic pathway and the late gene pathway-ANS (anthocyanidin synthase) on the tissues [38], which permitted us to compare the expression profiles of the genes in our study.

RNA quality control for all tissues

The RNA concentration differed tremendously between tissue samples (Supplementary information 2). However, we extracted the RNA from two biological plant groups for each tissue sample. DNase treatment and purification were performed on these biological plants in replicates. The purified RNA was measured using the spectrophotometer. Accepted RNA readings obtained within A260/280 ratio and above 1.8-2.0 were considered for the study [39]. The RNA integrity was determined for all different samples. A gel electrophoresis was used to ascertain visible fragment band or degradation since the signal was not satisfactory and difficult to verify on the spectrophotometer. According to the results, the intensity of bands for (28S/18S rRNA-ribosomal RNA) in all tissues were clear and

Figure 1: RNA gel electrophoresis of *Clivia miniata* plant tissues.

noticeable (Figure 1). The control measures was tested and confirmed that the RNA samples were of good quality. Finally, the RNA quality results were further used to synthesis cDNA tissues.

PCR amplification analysis

Chalcone synthase and anthocyanidin synthase from *C. miniata* were putatively annotated as the four unigene *CmiCHS11996, CmiCHS43839, CmiCHS50130* and *CmiANS*, which were previously isolated and sequenced by Maleka et al. unpublished (Supplementary information 3). The results indicate a high sequence identity (72%-92%) with the corresponding genes from other plant species after the nucleotide Blast on (blast.ncbi.nlm.nih.gov/). Blast analysis revealed that the partial sequence of *CmiCHS11996/CmiCHS43839* exhibited high percentage sequence similarity with *Narcissus tazetta* (JN227883) up to 92%, while the others-81% identity with Mimulusaurantiacus (EU305683), Perillafrutescens (AB002815), Hypericumandrosaemum (AF315345) and Anthuriumandraeanum (DQ421809), then up 80% with Catharanthusroseus (AJ131813), Solenostemonscutellaioides (EF522149), Juglansnigra (X94995), Perillafrutescens (AB002582), Pogostemoncablin (KJ768876) and Vacciniumashei (AB694904). The obtained CmiCHS sequences also shared 79% identity with the following plants; Tarenayahassleriana (XM010553649), *Lonicera japonica* (JX068609), *Lilium hybrid* (AB715424), *Curcuma longa* (JN017186) and 78% for *Arabidopsis thaliana* (DQ062415). While CmiANS showed percentage similarity with Lycorischinensis-90% (KC131464.1), 79% in Punicagranatum (KF841619.1) and *Ipomoea horsfallia*, (GQ180934.1), 78% in *Rosa rugosa* (KP768081.1), 77% in *Fragaria ananassa* (JX134095.1), 76% inStrelitziareginae (KC484623.1), 72% in Liliumspeciosum (AB911314.1).

To isolate the *CHS* and ANS genes, first strand cDNA of C. miniata was synthesized by reverse transcription using an Oligo $(dt)_{18}$ primer. A gene-specific primer and CHS reverse amplification primers were designed to amplify the missing *CmiCHS* sequences. Thereafter, cDNA tissues were amplified and approximately 1000 bp (base pair) was isolated and sequenced. The obtained sequences were aligned and compared with the full-length sequence of *Narcissus tazetta,* Lycorischinensisand other plants. However, these partial sequences lacked approximately 222 bp (for *cmiCHS 43839*) and 240 bp (*cmiCHS 11996*) nucleotides from the predicted full-length sequence of 1173 bp (JN227883). While an 80 bp short to obtain a full-length cmiANS according to *in silico* prediction of 1063 bp (KC131464). The amplicon size fragments of the genes were similar to what had been determined during *in silico* analyses.

Characterization and sequence analysis of *cmiCHS* and *cmiANS* genes

The PCR products which resulted in the amplification of cDNA fragments for *cmiCHS* and *cmiANS* (Figure 2) were ran on a 2% Agarose gel electrophoresis. Further steps of using the "BioFlux DNA/ RNA Purification Kit"-manufacturer's protocol were performed successfully to specifically isolate the amplified DNA fragments. The size of the amplified PCR fragments for *CHS* and ANS in C. *miniata* var. miniata 'Teleurstelling' (orange flower) was similar to the expected size determined via *in silico* analysis. The PCR fragments were purified, sequenced and analyzed to determine the content of the amplified PCR products which resulted to be the nucleotide genes from C. *miniata.* The sequences obtained from the three genes were assembled to a set of contiguous (contig) sequences using Cap 3-PRABI-Doua online tool (doua.prabi.fr/software/cap3). However, the ORF region of the contig sequence lengths started from (933 bp) for *cmiCHS 11996*, (951 bp) for *cmiCHS 43839* and (982 bp) for cmiANS respectively,
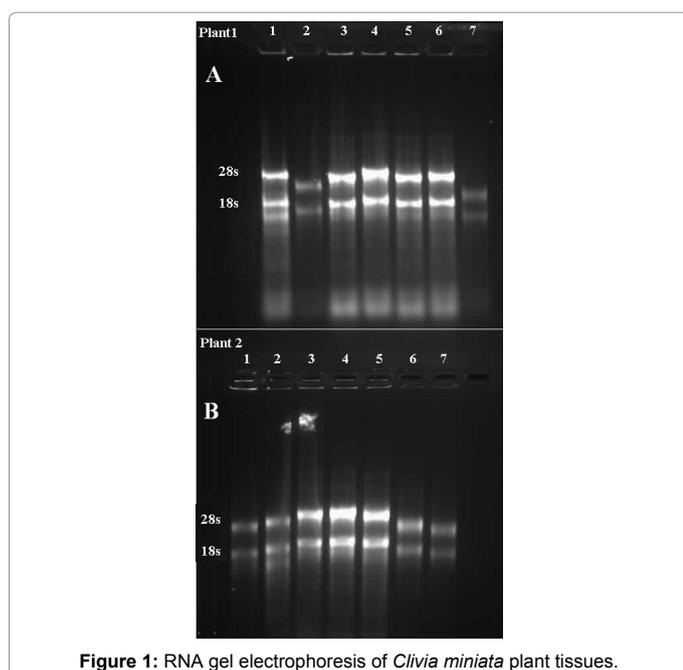
while the 3'-5'untranslated regions (UTR) were 47 bp, 154 bp, and 162 bp (Supplementary information 4a-4c). The electropherogram revealed single nucleotide polymorphisms (SNPs) of the genes; while the sequences were present in substantial quantity at certain positions (Supplementary information 5a-5c).

Analysis of *cmiCHS* and *cmiANS* predicted protein sequences and physiochemical properties

The amino acid sequences was deduced from the obtained contig cDNA sequences (expasy translate tool). The translated proteins had the following lengths respectively CM11996-276 a.a. and CM43839-316 a.a. However, the predicted putative molecular mass (Mw) and isoelectric point (pI) of the sequences were 30.5 kDa to 6.95 and 34.6 kDa to 7.54 kDa respectively. The CmiCHS protein sequences exhibited an extremely high sequence identity (>90%) to the other plant chalcone synthase's; more especially those plants from the Amaryllidaceae family. The highly homologous region of the CHS sequences across the species harnessed and identified important catalytically residues in the *CmiCHS* sequences; according to the ClustalO default tool parameter.

However, there are two important catalytic residues Cys[164], His[303] (highlighted with red background) out of the three known Cys[164], His[303], Asn[336] and two highly conserved (Phe) amino acids residue Phe[215] and Phe[265] in the sequences (highlighted with green background) that were found after multiple alignment of CmiCHS sequences with other similar CHS protein sequences (Figure 3). The predicted formulas for the two proteins are $C_{1351}H_{2182}N_{368}O_{395}S_{19}$ and $C_{1537}H_{2476}N_{418}O_{449}S_{20}$, while both

proteins were revealed to be unstable proteins based on their instability index prediction. However, there were no signal peptide sequences identified, and a weak hydrophobic nature of the protein according to the hydrophobicity region analysis in the amino acid sequence of the two respective genes; this information speculates non-secretory and soluble proteins. The transmembrane analysis showed that the amino acid sequences were all located outside the transmembrane; therefore no strong transmembrane helix were found amongst the amino acid residues.

According to the search and functional analysis on GenBank CDD (Conserved Domain Data-base; www.biochem.ucl.ac.uk/bsm/cath/Gene3D/) indicated that CmiANS belong to the 2-oxyglutarate iron-dependent oxygenase (2OG-FeII-Oxy) superfamily, which was further characterized by the 2OG-FeII-Oxy homologous domain. Moreover, this family incorporates enzymes that catalyze the synthesis of a plant hormone ethylene by oxidative desaturation of 1- aminocyclopropane-1-carboxylate and the hydroxylation (ACC) and the desaturation stage in the formation of other plant hormones, colour pigmentation and metabolites such as anthocyanidins, gibberellins and flavones [40,41].

Multiple sequence alignment of CmiANS and other plant ANS amino acid sequences revealed active site positions given on the right side (Figure 4). Identical, conserved and semi-conserved amino acids in the column are indicated with the symbols "*", ":" and ".", respectively. Important functional conserved residues are highlighted with a coloured background: green-the three residues contain active sites of Histidine family; Yellow-Aspartate residue family; conserved region of an active site for Arginine residue. These ANS sequences similarities, structural domains amongst the plants and the other 2OG-dependent dioxygases imply close structural similarities and functions.

Three-dimensional model of *cmiCHS*

The homology-based modeling of a 3D protein structure of *cmiCHS11996* and *cmiCHS43839* was constructed using YASARA HOMOLOGY MODEL (Supplementary information 6-URL link) for the chosen modelling parameters. Considering the fact that the target sequences were the only available information, possible templates were identified by 3 PSI-BLAST iterations (Position Specific Iteration Blast) extracted a position scoring matrix (PSSM) from Uni. Ref. 90, and then searching the PDB (Protein Data Base) for a match (i.e. hits with an E-value below the homology modelling cut-off 0.5). There were six (11996)/eight (cm43839) templates of protein sequences scored with the two obtained protein.

The six models were ranked and sorted according to their overall quality Z-scores. YASARA-MODEL combined the best parts of the six models to obtain a ranking hybrid model, which would increase the accuracy beyond each of the donors. The following sequence fragments were copied from other models; this indicates the first transfer during modelling was considered the most suitable for hybridization - in addition, the scores in the right column. The resulting quality of hybrid model Z-scores obtained is as follows for CM11996 protein-Satisfactory Packing 3D (-1.037) and an overall Quality Z-score of (-0.439), while for CM43839 revealed a Satisfactory Packing 3D (-1.113) and an overall Quality Z-score of (-0.686) respectively [Supplementary information 6-URL link].

The protein structure modelling showed *cmiCHS* monomer consisted of two structural domains (N-terminus and C-terminus) and secondary structures (α Helix, β sheet and Coils) were obtained to aid in the modelling process of the protein structure (Figure 5). However, there was Cys164 residue acting as an active site nucleophile for both structures; indicated in red and yellow colour. The two hybrid model
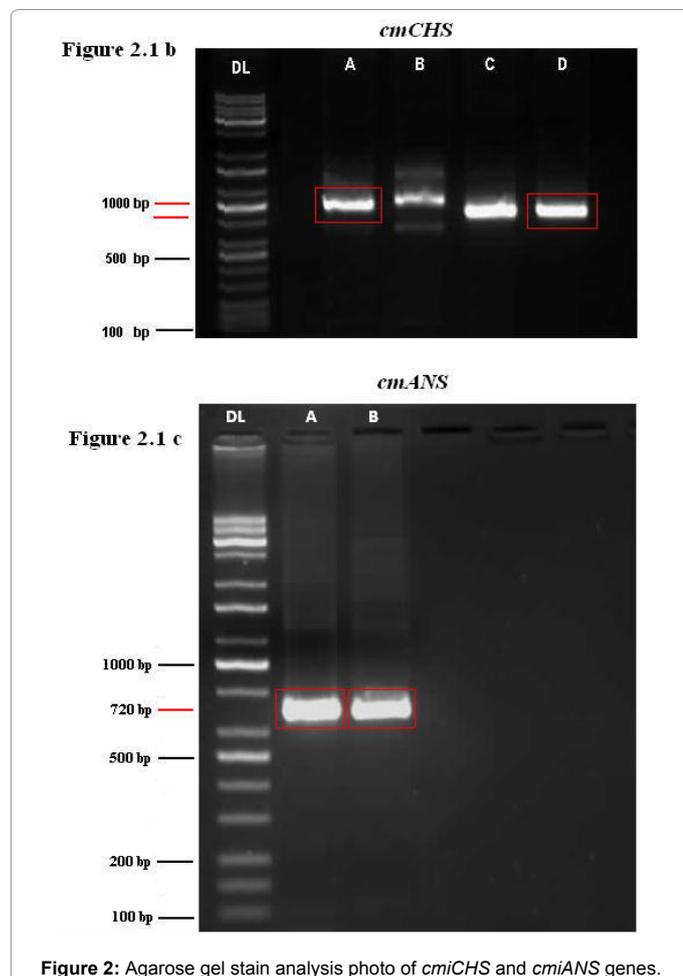

Figure 2: Agarose gel stain analysis photo of *cmiCHS* and *cmiANS* genes.

```
V.vinifera     MVSVGEIRKSQRAEGPATVLAIGTATPANCVYQADYPDYYFRITNSEHMTELKEKFKRMC    60
M.notabilis    MVTVEEVRKAQRAEGPATIMAIGTATPPPNCIDQSSYPDYYFRITNSEHKTELKEKFKRMC   60
G.hirsutum     MVTVEEVRKAQRAEGPATVLAIGTSTPPNCVDQSTYPDYYFRITNSEHKTELKEKFKRMC   60
D.cambodiana   MVAIDEIRRAQRADGPAAILAIGTANPSNCVEQSTYADYYFRVTNSEDKVELKEKFKRMC   60
N.tazetta      MVTIDEIRKAQRADGPATILAIGTANPQNCIEQSTYPDYYFRVTNSEHMVELKQKFKRMC   60
cmiCHS 11996   MVSIDEIRKMQRAEGPATILAIGTANPLNCIEQSAYPDYYFRVTNSEHMIELKQKFKRMC   60
cmiCHS 43839   MVSIIEIRKAQRAEGPATILAIGTANPLNCIEQSAYPDYYFRVTNSEHMIELKQKFKRMC   60
               **:: *:*: ***:***::::****:.* **: *: * *****:****.  ***:******

V.vinifera     EKSMINKRYMHLTEEILKENPNVCAYMAPSLDARQDMVVVEVPKLGKEAAVKAIKEWGQP   120
M.notabilis    EKSQIKKRYMYLHEEILKENPNICAYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQP   120
G.hirsutum     EKSMIKKRYMYLTEEILKENPNVCEYMAPSLDARQDMVVVEVPRLGKEAATKAIKEWGQP   120
D.cambodiana   DKSMIKKRYMYLTEEILKQNPNVCAYMAPSLDARQDMVVVEVPRLGKEAATKAIKEWGQP   120
N.tazetta      EKSMIKKRYMYLTEEILKQNPNLCAYMAPSLDARQDMVVVEVPRLGKAAVKAIKEWGQP   120
cmiCHS 11996   EKSMIKKRYMYLTEEILKQNPNVCAYMAPSLDARQDMVVVEVPRLGKEAAVKAIKEWGQP   120
cmiCHS 43839   EKSMIKKRYMYLTEEILKQNANVCAYMAPSLDARQDMVVVEVPRLGKEAAVKAIKEWGQP   120
               :** *:****:* *****:* *:* *******************:*** **.*********

V.vinifera     KSKITHLVFCTTSGVDMPGADYQLTKLLGLKPSVKRLMMYQQGCFAGGTVLRLAKDLAEN  180
M.notabilis    KSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAEN  180
G.hirsutum     KSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRVAKDLAEN  180
D.cambodiana   KSKITHLVFCTTSGVDMPGADYQLTKLLGLRPCVKRLMMYQQGCFAGGTVLRLAKDLAEN  180
N.tazetta      KSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAEN  180
cmiCHS 11996   KSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAEN  180
cmiCHS 43839   KSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRLAKDLAEN  180
               ****************************:*.*********************:*******

V.vinifera     NAGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAIIIGADPDTKIERPLFELV  240
M.notabilis    NKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAMIIGADPIPEVERPIFQLV  240
G.hirsutum     NKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAVIVGADPLSEIEKPMFELV  240
D.cambodiana   NRGARVLVVCSEITVVTFRGPSDTHLDSLVGQALFGDGAAAMIIGSDPIDTIEHPIFEIV  240
N.tazetta      NRGARVLVVCSEITAVTFRGPTDTHLDSLVGQALFGDGAAAMIIGADPVENVERPIFKLV  240
cmiCHS 11996   NRGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAMIIGADPVENVERPIFELV  240
cmiCHS 43839   NRGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAMIIGADPVENVERPIFELV  240
               *  ************.******:***********************:*:*:**   :*:*:*:*::*

V.vinifera     SAAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLVEAFKPIGISDWNSLFW  300
M.notabilis    SAAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIDKSLTEAFKPLGISDWNSIFW  300
G.hirsutum     SAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLAEAFQPLGISDWNSLFW  300
D.cambodiana   SAAQTLCPDSECAIEGHLREVGLTFHLLTDVPGIISKNIEKCLDDAFKPLGISDWNSLFW  300
N.tazetta      SAAQTLCPDSEGAIDGHLREVGLTFHLLKDVPGIISKNIERCLDDAFKPLGISDWNSLFW  300
cmiCHS 11996   SAAQTLCPDSEGAIDGHLREVGLTFHLLKDVPVFLL------------------------  276
cmiCHS 43839   SAAQTLCPDSEGAIDGHLREVGLTFHLLKDVPGIISKNIEKCLDDAFQPFIYQIGTRCSG  300
               *****: ***: **:*************.*** ::

V.vinifera     IAEPGGPAILDQVELKLGLKEEKLRATRHVLSEYGNMSSACVLFILDEMRKKSIEEGKGT  360
M.notabilis    IAEPGGPAILDQVEAKCGLKPEKLRATRHVLSEYGNMSSACVVFILDEMRRKGAEDGLKT  360
G.hirsutum     IAEPGGPAILDQVEAKLALKPEKLRATRHVLSEYGNMSSACVLFILDEMRKKSKEDGLGT  360
D.cambodiana   IAEPGGPAILDQVEAKLSLKPEKLRATRQVLSEYGNMSSACVLFILDEMRKKSAEEGKGT  360
N.tazetta      IAEPGGPAILDQVEEKLKLKEEKMRATREVLSEYGNMSSACVVFILDEMRKRSAAQGKGT  360
cmiCHS 11996   ------------------------------------------------------------  276
cmiCHS 43839   VAEPGGPAILITLSYP--------------------------------------------  316

V.vinifera     TGEGLEWGVLFGFGPGLTVETVVLHSLATQSTH    393
M.notabilis    TGEGLEWGVLFGFGPGLTVETVVLHSVAI----   389
G.hirsutum     TGEGLEWGVLFGFGPGLTVETVVLHSISA----   389
D.cambodiana   TAEGLKWGVLFGFGPGLTVETVALRSVATN---   390
N.tazetta      TGEGLEWGVLFGFGPGLTVETVVLHSVATS---   390
cmiCHS 11996   ---------------------------------   276
cmiCHS 43839   ---------------------------------   316
```

Figure 3: Alignment and active site regions in CHS proteins from different plants.

```
Camellia    ---MVA--TVAGIRVESLASSGVESIPKEYIRPQEELTSIGNVFEEEKK--EEGPQVPTV 60
Solanum     ---MVSEVVPTPSRVESLAKSGIQVIPKEYVRPQEELNGIGNIFEDEKK--DEGPQVPTI 60
Tulipa      --------MPTSARVESLSDSGLATIPKEYVRPESERDNLGDAFDEATKLDSAGPQVPVV 60
Allium      MTIESVIIAPPAPRVETLSKSNLHSIPLEYIRPEHERACLGDALEQLHN-SNSGPQIPIT 60
Lycoris     -----------MVRVESLASSGLDAIPSEYVRPECERDHLGDALEEVKK-AEKGPQIPII 60
CmiANS      -----------MVRVESLVSSDLNVIPSEYVRPECERDHLGDVLEEVKK-AEEGPQIPIV 60
                        ***:* .* :  ** **:**: *    :*: ::: :  . ***:*

Camellia    DLKDLVAE--DKEVRERCREALKKAATEWGVMHLVNHGIPDELMERVKAAGEGFFNQPVE 120
Solanum     NLKEIDSE--DKEIREKCHQELKKAAVEWGVMHLVNHGISDELIDRVKVAGGTFFDLPVE 120
Tulipa      DLAGFDST--DEKERAKCVEALRKAAEDWGVMHIVNHGIAKEVIEKVREVGKAFFDLPVG 120
Allium      DLD-----------SSDCIEKVTKAAKEWGVMHIVNHGISSELMEKVRAAGKAFFNLPLE 120
Lycoris     DLKGFDEESSDDIKRIKCIESVKIAAKEWGVMHITNHGISQELIEKVRAVGKGFFDLPME 120
CmiANS      DLKGFDEESSDDVKRIKCIKSVKIAAKEWGVMHITNHGISQELIEKVRAVGKGFFDLPME 120
            :*              * :  :  ** :*****:.**** .*::::*: .*   **: *:

Camellia    EKEKYANDHDSGNIQGYGSKLANNASGQLEWEDYFFHLVFPEDKRDMSIWPKTPSDYIPA 180
Solanum     EKEKYANDQASGNVQGYGSKLANSACGQLEWEDYFFHCVFPEDKRDLAIWPKTPADYIPA 180
Tulipa      EKEKYANDQESGDIQGYGSKLANNECGQLEWQDYFFHLIFPEEKTNLALWPKQPAEYTEV 180
Allium      AKEEYANDQSKGKIQGYGSKLANNASGQLEWEDYFFHLIFPDDKVDLSVWPKQPSDYIEI 180
Lycoris     MKEQYANDQSEGKIQGYGSKLANNSCGKLEWEDYFFHLIFPSDKVDMSIWPKQPSEYIEV 180
CmiANS      TKEQYANDQSEGKIQGYGSKLANNSCGQLEWEDYFFHLIFPSDKVDMSIWPKQPTDYIEV 180
             **:****: .*.:*********. .*:***:***** :**.:* :::::*** *::*

Camellia    TSEYAKQLRGLASKVLSALSLGLGLE-EGRLEKEVGGMEELHLQMKINYYPKCPQPELAL 240
Solanum     TSEYAKQIRNLATKIFAVLSIGLGLE-EGRLEKEVGGMEDLLLQMKINYYPKCPQPELAL 240
Tulipa      TKEFAKQLRVVATKMLSMLSLGLGLE-SGKLEKELGGMEELLMQMKINYYPKCPQPELAL 240
Allium      MQEFGSQLRILASKMLSILSLGLQLPTKDRLEQELKGPEDLLLQLKINYYPKCPQPHLAL 240
Lycoris     MQEFARQLRVVVSKMLAILSLGLGLKDEGKVETELGGMEDLLLQMKINYYPKCPQPDLAV 240
CmiANS      MQEFAEQLRVVASKMLAILSLGLGLKDEGKLETGLGGMEDLLLQMKINYYPKCPQPDLAL 240
            .*:. *:* :..:*::: **:** *   . ::*   :  * *:* :*:*********.**:

Camellia    GVEAHTDVSALTFILHNMVPGLQLFYEGKWVTAKCVPNSIIMHIGDTVEILSNRKYKSIL 300
Solanum     GVEAHTDVSALTFILHNMVPGLQLFYEGKWVTAKCVPNSIIMHIGDTIEILSNGKYKSIL 300
Tulipa      GVEAHTDVSSLTFLLTNMVPGLQLYYGDKWVIAECVPDSLLVHIGDTLEILSNGSYRSIL 300
Allium      GVEAHTDVSALSFILHNNVPGLQVLYEGEWVTAKLVPDSLIVHVGDSLEILSNGIYKSVL 300
Lycoris     GVEAHTDVSALSFILHNNVPGLQVFYDDKWVSAQLVPDSIIVHVGDALEILSNGMYKSVL 300
CmiANS      GVEAHTDVSALSFILHNNVPGLQVFYNDKWVSAQLVPDSIIVHVGDTLEILSNGMYKSVL 300
            *********:*:*:* * *****: *   :** *: **:*::*:*:**::***** *:*:*

Camellia    HRGLVNKEKVRISWAVFCEPPKEKIILQPLPETVTEEEPPLFPPRTFAQHIQHKLFRKTQ 360
Solanum     HRGVVNKEKVRISWAIFCEPPKEKIMLKPLPETVTEAEPSQFPPRTFAQHMAHKLFKKVD 360
Tulipa      HRSLVNKDRVRISWAVFCEPPKETIVLQPLPELVSEAAPAKFPPRTFKQHIQHKLFKKTE 360
Allium      HRGLVNKEKVRISWAVFCEPPKDAVVLKPLDEVVTDDAPARYTPRTFAQHLERKLFKKKV 360
Lycoris     HRGLVNKEKVRISWAVFCEPPKDKILLRPLQELLTNEKPAKFTPRTFAQHLQRKLFKKTT 360
CmiANS      HRGLVNKEKVRISWAVFCEPPKDKIVLRPLQELVTDEKP-------------------- 360
            **.:***::*****:*****: ::*:** * :::  *

Camellia    VLGGK------------------------------------------------------ 420
Solanum     NDAAAEQKVFKKDDQDSAAVHKASEKDDRDIVAEHIVLKEDKQDSAVEQKAFKKVDQDVV 420
Tulipa      EELALPK---------------------------------------------------- 420
Allium      GDLDDSDV--------------------------------------------------- 420
Lycoris     EDFTPEM---------------------------------------------------- 420
CmiANS      ----------------------------------------------------------- 420
```

Figure 4: Alignment and active site regions in ANS proteins from different plants.

Figure 5: 3D structure of the *cmiCHS* protein obtained by YASARA model.
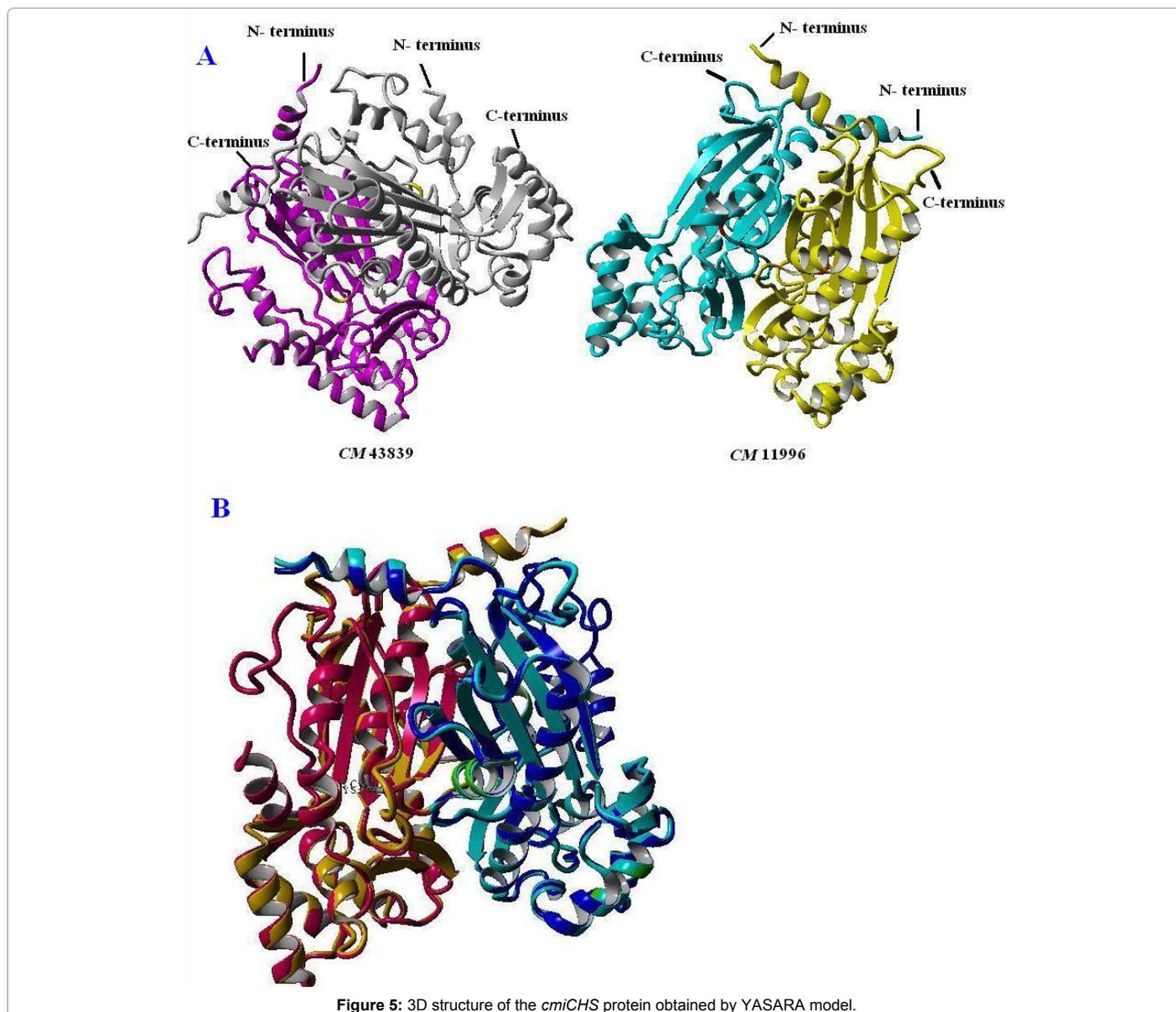
structures of the proteins were superimposed to compare the different protein structures based on either their C-alpha, residue or backbone atoms; and they both showed an identity on the residue number of their active sites (Cys164) and similar residue in both domains of two the proteins.

Phylogenetic analysis of chalcone synthase and anthocyanidin synthase

The phylogenetic tree of *CHS* and ANS plant protein sequences were constructed based on the high identity percentage (>90%) of amino acid sequences of the unigene (*cmiCHS11996*, *cmiCHS43839* and *cmiANS*) and other plant sequences. In order to characterize the evolutionary relationship of *cmiCHS* and *cmiANS* unigene, sequences were aligned using ClustalO Multiple alignment tool; and was further used to construct a neighbour-joining (NJ) phylogenetic tree with 1000 bootstrap replicates. The phylogenetic tree was constructed using MEGA software tool and Jones-Taylor-Thornton (JTT) model parameter was taken into consideration. The putative *CHS* and *ANS*

unigene of C. *miniata* were grouped according to identical plant gene sequences and while the other clade groups were from other plants (Figures 6 and 7).

The results found after the phylogenetic analysis are in agreement with the hypothetical proposition of the ancestry unigenes for the *CHS* or *ANS* superfamily. This suggest that the *cmiCHS* and *cmiANS* unigene proved to be informative and appeared as the recent common ancestor to all the plant *CHS* or *ANS* superfamily genes; and the distance percentile of those particular branch in the phylogenetic tree was relatively high (100%).

Amplification specificity

The cDNA samples were used to run a gradient PCR for all the genes and the orange stage 5 flower cDNA was reported as our reference temple (Figure 8). The amplification of DNA in cDNA tissues produce a false evaluation of the gene expression level of a gene or might even detect false expression. However, we did not observe non-specific binding after running the tissue samples on a gel. The Agarose gel

verified amplification fragments that was similar to the *in silico* length predicted before the PCR. A noticeable 20 bp DNA Ladder (Thermo Scientific) viewed on the left and a Universal DNA Ladder (Kapa Biosystem) loaded on the right part of the gel; were used to support each amplified length of the genes. The report on the gel (Figure 8) revealed noticeable bands and the fragment A1 (*Ef1α gene*), A2 (*18s gene*), A3 (*CmiANS*) and A4 (*CmiCHS 50130*). The temperature sequence are [(53°C - 63°C)-2°C range for the reference genes and *CmiANS*; (51°C - 61°C)-2°C apart for target gene *CmiCHS11996* and *43839*]. We choose (lane 5) for A1 and the amplicon length of (148 bp), A2 (130 bp) and A3 (105 bp) because it corresponded with the annealing temperature of (61°C), while A4 (lane 4-120 bp amplicon length) was the preferred
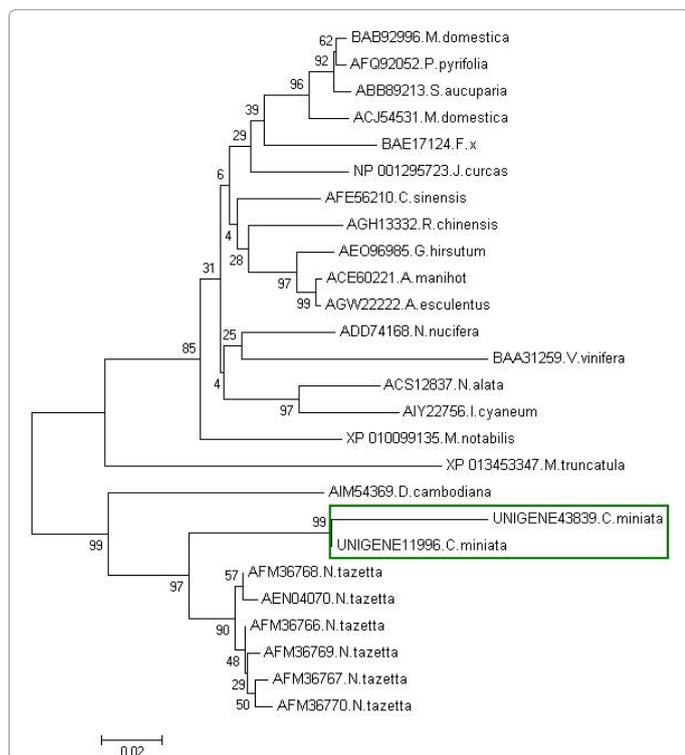
Figure 6: Neighbour-joining phylogenetic tree of CHS genes from different plants.
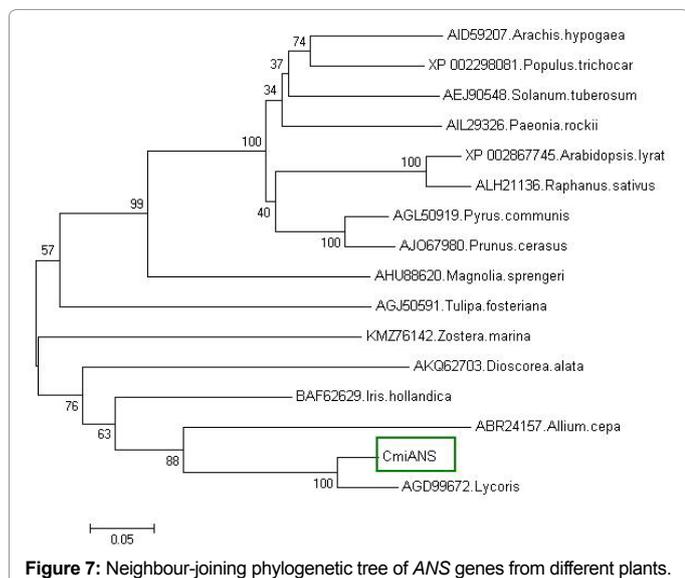
Figure 7: Neighbour-joining phylogenetic tree of *ANS* genes from different plants.
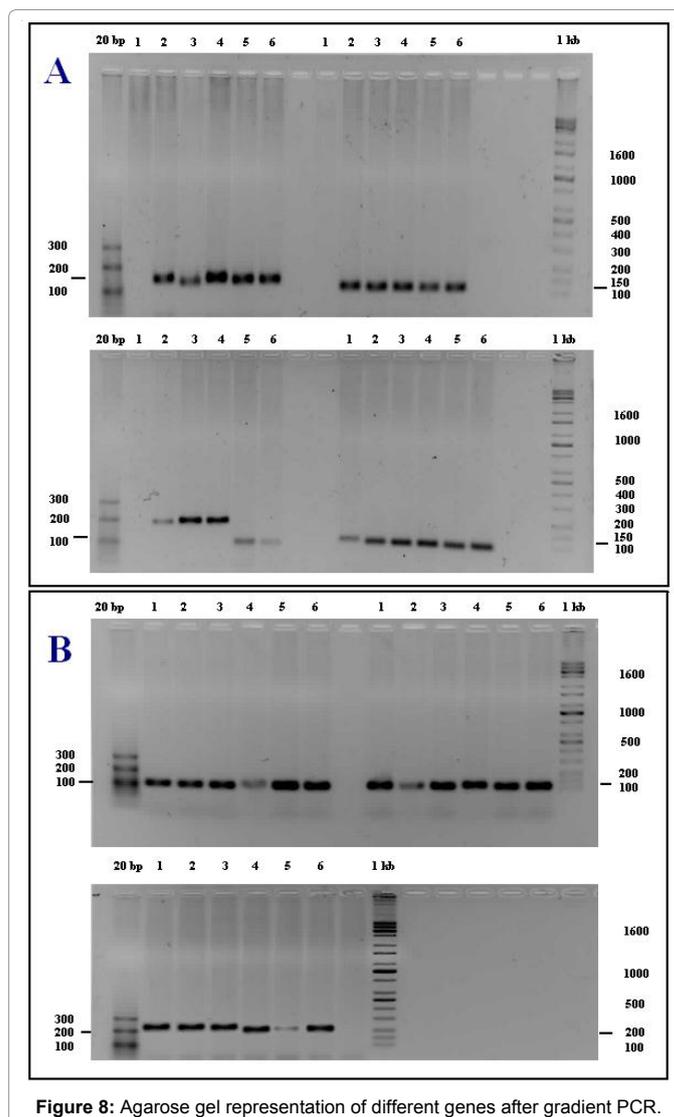
Figure 8: Agarose gel representation of different genes after gradient PCR.

lane for the target gene (57°C). The order for gel picture B is as follows; B1 (*CmiCHS 11996*)-lane 4 and 98 bp length, B2 (*CmiCHS 43839*)-lane 4 and 102 bp length and B3 (PP2A)-lane 5 and an amplicon length of 226 bp.

Reference gene selection and relative gene expression profile in different tissues

The possibility of normalizing genes is acceptable and these gene expression level assays should be normalized by critical conventional selection of more than two genes [37]. However, the three reference genes amplified in all the cDNA tissues were used for this test. The Ct Mean values were calculated and then imported to Normfinder Excel spread sheet. The data was used to log converted the Ct values which assisted in identifying the optimal normalization amongst the three reference genes [Supplementary information 7]. The analyzed data ranked PP2A as the best possible stable gene according to its expression stability for experimental assay [38-41].

In this study, the detection of amplicon increases using the fluorescent dye SYBR® Green ABI prism. While the PCR reactions advances in real-time with the primer set, the emission of a fluorescent signal was detected when the intercalating dye excites and emits
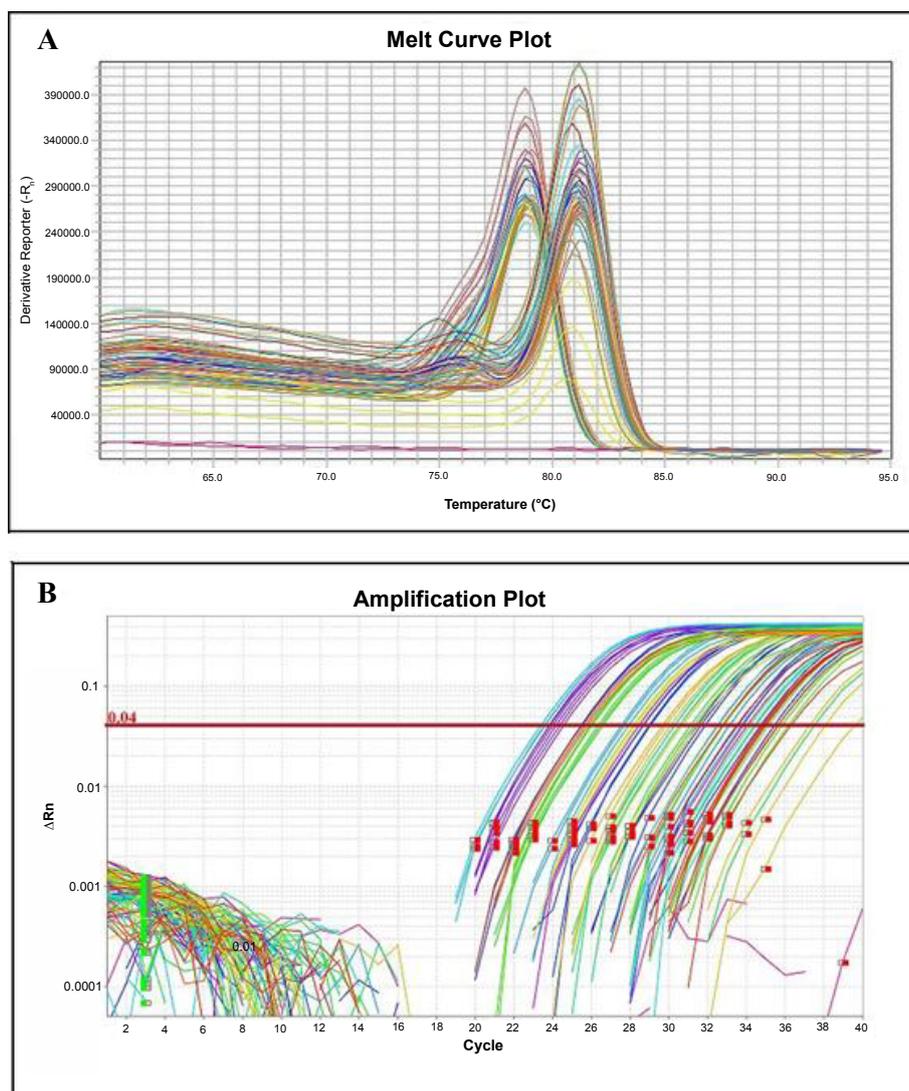
Figure 9: Graphical representation of the melt curve and amplification plot.

with the stranded DNA formed [42,43]. This fluorescent mechanism revealed an amplification plot and the melt curve of all the tissues after the gene expression assay. The amplification showed visible ΔRn (Rn-normalization reporter) values per cycle, the flower tissues emits and amplifies faster than the other tissues which happen to amplify late. While the melt curve revealed the distinct curves between the primer dimers and the melt curve of the amplified tissues, however the NTC was constant (Figure 9).

The comparative threshold (Ct) method (ΔΔCt) was used to analyze gene expression of CmiCHS and CmiANS in all the tissues of the two biological plants of *C. miniata*. This was determined by normalizing the reference gene and the relative amount of each Ct values for the target genes, obtaining the average Ct values of plant 1 (P1) and plant 2 (P2) in comparative to a calibrator by calculating $2^{-\Delta\Delta Ct}$ [Supplementary information 8 and 9]. The tissue which had the highest (ΔCt value) was assigned as the calibrator. The histogram of the relative expression profile of the structural genes revealed higher differentially expressed levels of the flowering tissues than in the leaf, scape, style and stigma-but slightly high in the berries (c). The highest expression level of *CmiCHS 11996*, *CmiANS*, *CmiCHS 43839* and *CmiCHS 50130* was observed in the tepals (orange and yellow flowers)

which were respectively 3.10- (O1); 3.62- (O5), and 3.10- (Y5) folds as compared to lower expression levels in leaves, style, stigma and scape. In addition, the berries showed slight higher expression levels when compared to the non-colour pigmentation tissues but lower to the flowering tissues. However, yellow flower- developmental stage 1 was used as our calibrator for the analysis. It was noted that *CmiCHS 43839* and *CmiCHS 50130* mRNA was low in leaf, style and stigma and scape throughout the flower development and they showed no significant difference. While similar pattern was observed in these genes (*CmiCHS 11996* and *43839*; *CmiCHS 50130* and *CmiANS*) respectively exhibited no significant difference of expression level in the orange flower-developmental stage 1 (Figure 10).

Discussion

Since chalcone synthase (*CHS*) is core enzyme in flavonoid pathway which encodes gene family members, research revealed that each gene family was identified based on the gene duplication and a functional divergence driven by positive selection was observed [44]. Thus, CHS from other higher plants are encoded by multigene family and they have been documented in a number of studies (ref). Matsumura [45] identified nine CHS genes in Glycine max, and seven CHS genes in
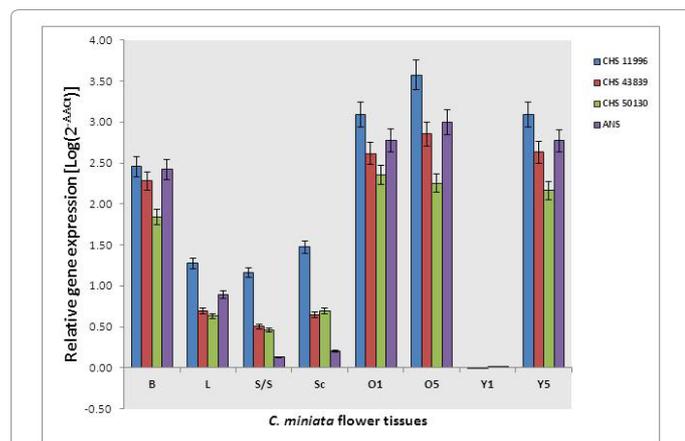
Figure 10: Histogram representation of *cmiCHS* and *cmiANS* gene expression.

Pisum sativum [46]. While Farzad [47] characterized multiple copies of CHS gene family in Viola cornuta. The fact that chalcone synthase is influential to Clivia flower pigmentation phenotype, it is vital to gain further information on the molecular activities of the CHS gene-family in Clivia. The molecular tools would assist in breeding or play significant role in modifying flower colours in other Clivia species. Molecular approaches were applied in-order to characterize *CmiCHS 11996*, *CmiCHS 43839* and *CmiANS* gene family in Clivia. In this study, we synthesized cDNA from RNA from C. *miniata*. The unigenes of *CmiCHS 11996*, *CmiCHS 43839* and *CmiANS* were successfully amplified and a concrete sequences were obtained as described earlier in section 2.4.

Studies have shown that flavonoid biosynthetic genes share high sequence similarity among different monocot/dicot plant species regarding the phylogenetic characterization [48,49]. Hence, the primer set designed from a monocot sequence data of *Narcissus tazetta* CHS was suitable to amplify fragments of *C. miniata* CHS gene since Clivia are monocots and also similar in terms of phylogenetic taxa. The *in silico* analysis of CHS extracted from the *Narcissus tazetta* var. chinensis cultivar Baihua sequence revealed appropriate regions for designing primers. This CHS sequence of *N. tazetta* showed a highly similar percentage of 92% after a BLAST search of chalcone synthase gene in the Gene Data Base (NCBI). Designated sets of GSP primers and CHS forward/reverse primers were annotated as putative CmiCHS from *N. tazetta* ESTs. The PCR that was performed with these primers using cDNA template synthesized from an RNA sample of *C. miniata* amplified fragments of *CmiCHS*. The PCR amplification was successful and the results revealed the expected length of 1000 bp for *CmiCHS* genes (lane A - lane D) and 720 bp of *CmiANS* gene. In most cases, the PCR products correlate to the expected size and subsequent analysis confirms the identity of the PCR products by performing subsequent sequence analysis. In accordance to the electropherogram of *CmiCHS* sequences, it revealed single nucleotide polymorphisms (SNPs) at certain positions suggesting the presence of heterozygous CHS alleles expressed in the flower tissues and also mismatch of nucleotides due to mechanical error [Supplementary information 5a-5c]. The obtained sequences contain an open reading frame (ORF) 933, 951, 983 bp (*cmiCHS 11996-43839* and *CmiANS*) respectively. The translated amino acid sequences (309, 316, and 327) residue corresponded to the deduced protein sequence of both CHS and ANS plants predicted during *in silico* analysis.

Amino acid sequence identity between the obtained genes and other CHS or ANS super-family genes were closely related and the genes varied in-terms of percentage similarity. The high degree of sequence identity (>90%) among these genes strongly indicates that the obtained genes is encoded by the enzymes - CHS and ANS [48-50]. The contig of both gene sequences did represent the overlaps of the sequences obtained during sequence analysis which assisted in determining substantial lengths. However, the contig showed a high percentage similarity of CHS and ANS genes from other plant protein sequences submitted at NCBI gene database. This suggest that the unigenes contains homologous amino acid sequences of domain that comprises of amino acid residues and they contain the active site, product binding site, polypeptide binding site and malonyl-CoA binding site [21,51]. The InterPro Scan analysis, the domain of anthocyanidin synthase of C. *miniata* contains similar characteristic featuring Non-haem-dioxygenase N-terminal (48-60 amino acid fragment), Isopenicillin N synthase like region (161-209) and 2OG-FeII_Oxy (210-308 amino acid fragment) super family in correspondence to other plant members. The CmiANS possess the active sites of His (H_{97}, $H_{245, 283}$) and Asp (D_{247}) residues. These residues are believed to regulate iron at the catalytic center of the iron-containing oxygenases and the 2-oxoglutarate-dependent enzymes. In addition, Wilmouth [52] revealed that the residue label in red (Arg-R_{302}) are conversed in ANS, which is supposed to contribute to the binding of 2-oxyglutatrate and which may also provide positive charge. The chalcone/stilbene synthase domain according to the result analyzed by InterPro Scan (IPR00199) indicated the N-terminal domain (PF00195) started from amino acid 6 to 228, while the C-terminal domain (PF02797) started from 238 to 312 amino acids from 316 *CmiCHS 43839* protein sequence. Then from the 276 *CmiCHS 11996* protein sequence submitted, the N-terminal domain started from 5 to 228, and then from 238 to 273 for the C-terminal of protein sequences submitted; thus, in comparison to Chalcone/stilbene synthase protein match after the InterPro functional analysis. Ferrer [53] reported that the domains of chalcone synthase are structurally similar to domains in thiolase and beta-ketoacyl synthase while the differences in activity are noticed in the N-terminal domain. The chalcone synthase-like protein family has an E-value of 2.52e-09 for the domain selected and it's actually a conditional score on the domain being a Thiolase-like superfamily member. While the active site contains two important homologous catalytic triad of Cys^{164} and His^{303} highlighted in colour 'red', while Asn^{336} residue was not obtained in the sequence because of its incomplete amino acid residue 3' flanking region. Cys^{164} residue plays role as the active-site nucleophile in polyketide synthesis, and makes known the importance of His^{303} and Asn^{336} in the malonyl-CoA decarboxylation reaction. These two proteins from CmiCHS11996 and CmiCHS43839 also revealed two vital amino acids of Phe^{215} and Phe^{265} highlighted in 'green'. According to Jez [54], the Phe^{215} and Phe^{265} act as "gatekeepers" to stop the lower proteins. This mechanism reduces the access of water to the active while housing substrates and intermediates of various shapes and sizes of other proteins. Phe215 residue also functions to orient substrates at the active site during elongation of the polyketide intermediate. Therefore, these three amino acids (Cys^{164}, Phe^{215} and His^{303}) are found at the intersection of the CoA-binding tunnel and the active site cavity; they play essential and distinct role during malonyl-CoA decarboxylation and chalcone formation. In addition, the phylogenetic analysis demonstrates that our CmiCHS Unigene belong to the corresponding clusters (grouped into monocot and; which is from the same CHS superfamily of Amaryllidoideae. This observation correlates with previous study of [55].

According to Hooft [56], the total BLAST alignment score revealed a quality score in the PDBFinder2; ranging from 0.000 (terrible) to 1.000 (perfect). The secondary structure was obtained and did predict for the target sequence; these helped in alignment correction and loop

modelling. The secondary structure was successfully determined by running the *CmiCHS 11996* and *CmiCHS 43839* proteins in the PSI_BLAST to create a target sequence profile, then feeding it to the PSI-Pred secondary structure prediction algorithm [57]. In the alignment, 313 of 316 target residues (99.1%) aligned to template residues. In the midst of the aligned residues, sequences revealed 79.9% identity and 87.9% similarity (BLOSUM62 score is >0). Once the side-chains had been formed, optimized and fine-tuned during modelling, the newly formed parts were subjected to a stimulated annealing minimization. However, this means the backbone atoms of the residues alignment were kept intact to prevent potential damage. The 3D structure model was generated by YASARA Structure Version 15.7.12 (URL-Supplementary information 6).

According to the 3D structure formed, YASARA collectively modelled the best part of the 8 protein models obtaining a hybrid model which increased the accuracy beyond each of the protein contributors. Each fragment was copied from other models where by the last copy are basically the initial model considered most suitable for hybridization and which the scores revealed an accepted Z-scores -0.686 for *CmiCHS 43839* and -0.439 *CmiCHS 11996* respectively (Supplementary information 6). Some measures were taken into consideration like for instance; if the hybrid model scores was worse than the model from which it was initially derived, this is not a bad result since the hybrid model often covers more residues. The Z-score explains how far the standard deviations the model quality is further from the average high-resolution X-ray structure. The total Z-scores for all the models was calculated as the weighted averages of the individual Z-scores using the formula (Overall Z-score=0.145-Dihedrals+0.390-Packing1D+0.465-Packing 3D). The overall score entails the correctness of backbone - (Ramachandran plot), the side-chain dihedrals and also for the packing interactions.

The putative structural genes of interest can be classified into two groups regarding their position in the pathway and the expression regulation in the early biosynthetic gene CHS, while ANS is a late biosynthetic gene in respect to our study [58]. Previous study done by Vandesompele [59] reported misinterpretation of results due to inconsistent concentration, quality of RNA and retro-transcription efficiencies during synthesis of cDNA. In the present study, optimization procedures were performed during gene expression which improved the quality of cDNA. Biological replicates were taken into consideration and assembled for different C. miniata tissues, which is a better approach to avoid the physiological condition of one plant influencing the overall gene expression. Hence, biological replicates are prerequisite to successful gene expression [60]. The quality of RNA is remarkable for gene expression assay. We purified the RNA samples and determined the absorbance concentration. Though, looking only at the absorbance ratio/concentration can lead us to wrong assumptions regarding the RNA quality. Nonetheless, we obtained visible optimal 28S/18S rRNA fragments of ratio of 2.1 for (A_{260}/A_{280}) after the quality gel-electrophoresis assessment. Furthermore, the cDNA for all tissues were tested by performing PCR amplification with reference genes, target genes and NTC. The fact that the successful report on RNA concentration [Supplementary information 2] showed no tissue contamination, thus which strengthens our study. Nonetheless, several quantification strategies with standardizing techniques are accessible, but based on the PCR efficiencies (E) for their calculations [33].

In this study, we demonstrated the comparison of four gene transcript levels by qRT-PCR in different *C. miniata* tissues which are involved in colour pigmentation during flavonoid biosynthesis. The early biosynthetic genes *CmiCHS 11996*, 43839 and 50130 and the late

gene CmiANS showed tremendous abundance of expression in the flower tissues (O1; O5; Y5) and partially in the berries as compared to the leaves, style and stigma and scape (Supplementary information 8 and Figure 10). In a study performed by [36], revealed the higher production of CHS and DFR expression in the orange flowers of *C. miniata* compared to the yellow flower cultivar. In respect of transcript level of *CmiCHS 11996*, which had the highest fold number of 3.62-fold in (O5 tissue) is significantly different from other tissues. This explains the up-regulation of the gene in a fully developmental stage (full flower blossom). The target genes in the (O1, O5 and Y5) showed similarity trend of expression, except of *CmiCHS 50130* in (Y5). Though, the depicted graph shows no significant difference between *CmiCHS 11996*, *43839* and *CmiANS* genes in (B) and *CmiCHS 50130* gene in (O1; O5; Y5) in respect to an overlap of their error bars. In the leaves, style, stigma and scape; *CmiCHS 43839*, *50130* and *CmiANS* showed similar trend in their expression with lower levels of expression profiles, (0.20-fold) in style and stigma being the lowest expression level compared to the other flowering tissues.

The expression of *CmiCHS* and *CmiANS* genes was observed to be generally higher in all the flowering tissues including berries than in rest of the tissues. These findings support the theory of abundant accumulation of anthocyanin in the flowering tissues being a secondary metabolite responsible for pigmentation. Though, the correlation of between these putative structural genes and it accumulation of anthocyanins could be attributed to some other secondary metabolites (alkaloids) in respect to some tissues (leaves and roots) in Clivia been a higher plant [61]. The result of this study revealed an increase of *CmiCHS* throughout the flowering stage; while the total anthocyanin in tepals (orange flower-developmental stage 5) was found to be higher. Chalcone synthase being an early enzyme partaking in the synthesis of anthocyanin, flavones and flavonols [62] and though the entire flavone and flavonols content shows no significant difference between the orange and yellow flower (tepals) and the other tissues. Similar expression levels of *CmiANS* were continuously and significantly higher in the tepals; which implied more accumulation anthocyanin. Since *CmiANS* is late biosynthetic gene which only participates in the formation of anthocyanin, its expression pattern presents a significant correlation with anthocyanin accumulation. This result was similar and in accordance to a report on petunia [63,64].

Conclusion

The most vital part for manipulating Clivia flower colour was to comprehend information of the key enzymes in anthocyanin biosynthetic pathway. Molecular tools aided in performing PCR procedure to amplify CmiCHS and CmiANS cDNA fragments, obtaining a successful amino acid sequences. However, the partial sequences obtained lack some missing nucleotides; but the important motif identified from the highly conserved region of the protein sequences entails the proteins are actively involved in flavonoid biosynthetic pathway. These results correlate to the literature on CHS and ANS plants which provide meaningful information for studying *CHS* gene in Clivia plants. However, all the putative structural genes in this study were expressed distinctively between the tepals (orange and yellow flowers) and other tissues resulting to an expected higher production of anthocyanin. There were more abundant expression of both early and late biosynthetic genes (*CmiCHS* and *CmiANS*) detected in the floral tissues than in scape, style and stigma and leaves. The temporal expression profiles of these putative structural genes suggest a strong correlation with the production of colour pigments in *C. miniata*. This result obtained from the analyzed data finally suggest that temporal regulation of *CmiCHS* and *CmiANS* expressions could

directly change the amount of anthocyanin accumulation in tepals of *C. miniata*, which would assist and impact the horticultural industry and further research.

Acknowledgements

References

1. Meerow AW, Fay MF, Guy CL, Li QB, Zaman FQ, et al. (1999) Systematics of Amaryllidaceae based on cladistic analysis of plastid sequence data. Am J Bot 86: 1325-1345.

2. Lindley J (1828) Clivia nobilis. Edwards's Botanical Register 14: 1182.

3. Regel E (1864) Clivia miniata Lindl. Amaryllideae. Gartenflora. 14: 131.

4. Hooker W (1856) Clivia gardenia. Curtis's Botanical Magazine.

5. Dyer R (1943) Clivia caulescens The Flowering Plants of South Africa.

6. Rourke JP (2002) Clivia mirabilis (Amaryllidaceae: Haemantheae) a new species from Northern Cape, South Africa. Bothalia 32: 1-7.

7. Murray BG, Ran Y, De Lange fls PJ, Hammett KRW, Truter JT, et al. (2004) A new species of Clivia (Amaryllidaceae) endemic to the Pondoland Centre of Endemism, South Africa. Bot J Linn Soc 146: 369-374.

8. Ran Y, Simpson S (2005) In vitro propagation of the genus Clivia, Plant Cell. Tissue and Organ Culture 81: 239-242.

9. Chandler SF, Brugliera F (2011) Genetic modification in floriculture. Biotechnol Lett 33: 207-214.

10. Maleka MF, Albertyn J, Spies JJ (2013) The floriculture industry and flower pigmentation - a review. Philos Trans Genet 2: 55-110.

11. Tanaka, Y, Sasaki, N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J cell Mol Biol 54: 733-749.

12. Zhou XW, Fan ZQ, Chen Y, Zhu YL, Li JY, et al. (2013) Functional analyses of a flavonol synthase-like gene from Camellia nitidissima reveal its roles in flavonoid metabolism during floral pigmentation. J Biosci 38: 593-604.

13. Kishimoto S, Sumitomo K, Yagi M, Nakayama M, Ohmiya A (2007) Three Routes to Orange Petal Color via Carotenoid Components in 9 Compositae Species, J Japane Soc Hort Sci 76: 250-257.

14. Ferrer JL, Austin MB, Stewart C Jr, Noel JP (2008) Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiol Biochem 46: 356-370.

15. Mol J, Grotewold E, Koes R (1998) How genes paint flowers and seeds. Trends Plant Sci 1385: 212-217.

16. Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 5: 218-223.

17. Buer CS, Imin N, Djordjevic MA (2010) Flavonoids: new roles for old molecules. J Integr Plant Biol 52: 98-111.

18. Falcone Ferreyra ML, Rius SP, Casati P (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci 3: 222.

19. Harborne JB (1994) The Flavonoids Advances in Research Since 1986. London: CRC Press.

20. Olsen KM, Hehn A, Jugda H, Slimestad R, Larbat R, et al. (2010) Identification and characterisation of CYP75A31, a new flavonoid 3'5'-hydroxylase, isolated from Solanum lycopersicum. BMC Plant Biol 10: 21.

21. Schrader J (1997) A family of plant-specific polyketide synthases: facts and predictions. Trends Plant Sci 2: 373-378.

22. Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Biol 57: 761-780.

23. Sommer H, Bonas U, Saedler H (1988) Transposon-induced alterations in the promoter region affect transcription of the chalcone synthase gene of Antirrhinum majus. Mol Gen Genet 211: 49-55.

24. Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. Proc Natl Acad Sci USA 96: 12929-12934.

25. Pang Y, Shen G, Wu W, Liu X, Lin J, et al. (2005) Characterization and expression of chalcone synthase gene from Ginkgo biloba. Plant Sci 168: 1525-1531.

26. Koes RE, Spelt CE, van den Elzen PJ, Mol JN (1989) Cloning and molecular characterization of the chalcone synthase multigene family of Petunia hybrida. Gene 81: 245-257.

27. Clegg MT, Durbin ML (2003) Tracing floral adaptations from ecology to molecules. Nat Rev Genet 4: 206-215.

28. Yuan Y, Ma X, Shi Y, Tang D (2013) Isolation and expression analysis of six putative structural genes involved in anthocyanin biosynthesis in Tulipa fosteriana. Scientia Horticulturae 153: 93-102.

29. Turnbull JJ, Sobey WJ, Aplin RT, Hassan A, Schofield CJ, et al. (2000) Are anthocyanidins the immediate products of anthocyanidin synthase?. Chem Commun 24: 2473-2474.

30. Turnbull JJ, Nakajima JI, Welford RWD, Yamazaki M, Saito K, et al. (2004) Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis: Anthocyanidin synthase, flavonol synthase, and flavanone 3-hydroxylase. J Biol Chem 279: 1206-1216.

31. Kim S, Binzel ML, Yoo KS, Park S, Pike LM (2004) Pink (P), a new locus responsible for a pink trait in onions (Allium cepa) resulting from natural mutations of anthocyanidin synthase. Mol Genet Genomics 272: 18-27.

32. Xu F, Cheng H, Cai R, Li LL, Chang J, et al. (2008) Molecular cloning and function analysis of an anthocyanidin synthase gene from Ginkgo biloba, and its expression in abiotic stress responses. Mol Cells 26: 536-547.

33. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408.

34. Phaffl W(2004) Quantification strategies in real-time PCR. 87-112.

35. Schmittgen TD (2006) Quantitative gene expression by real-time PCR: a complete protocol. New York: TefikDorak.

36. Viljoen CD, Snyman MC, Spies JJ (2013) Identification and expression analysis of chalcone synthase and dihydroflavonol 4-reductase in Clivia miniata. South African Journal of Botany 87: 18-21.

37. de Keyser E, Desmet L, Van Bockstaele E, De Riek J (2013) How to perform RT-qPCR accurately in plant species? A case study on flower colour gene expression in an azalea (Rhododendron simsii hybrids) mapping population. BMC Mol Biol 14: 13.

38. Cooman D, Everaert ESW, Facha P, Casteele KV, Sumere CFV (1993) Flavonoid biosynthesis in petals of Rhododendron simsii. Phytochemistry 33: 1419-1426.

39. Luis Oñate-Sánchez, Jesús Vicente-Carbajosa (2008) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. Biomed Central 7: 17.

40. Zhang Z, Barlow JN, Baldwin JE, Schofield CJ (1997) Metal-catalyzed oxidation and mutagenesis studies on the iron(II) binding site of 1-aminocyclopropane-1-carboxylate oxidase. Biochemistry 36: 15999-16007.

41. Lukacin R, Britsch L (1997) Identification of strictly conserved histidine and arginine residues as part of the active site in Petunia hybrida flavanone 3P-hydroxylase. Eur J Biochem 757: 748-757.

42. Shipley GL (2005) Real-Time Quantitative PCR: Theory and Practice, in Encyclopedia of molecular cell biology and molecular medicine (Ed). Wiley-VCH Verlag Weinheim 481: 523.

43. Shipley GL (2006) An Introduction to Real-Time PCR, in Real-time PCR. In: Dorak MT (Ed). New York: Taylor & Francis Group 1: 37.

44. Helariutta Y, Kotilainen M, Elomaa P, Kalkkinen N, Bremer K, et al. (1996) Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification. Proc Natl Acad Sci U S A 93: 9033-9038.

45. Matsumura H, Watanabe S, Harada K, Senda M, Akada S, et al. (2005) Molecular linkage mapping and phylogeny of the chalcone synthase multigene family in soybean. Theor Appl Genet 110: 1203-1209.

46. Ito M, Ichinose Y, Kato H, Shiraishi T, Yamada T (1997) Molecular evolution and functional relevance of the chalcone synthase genes of pea. Mol Gen Genet 255: 28-37.

47. Farzad M, Soria-Hernanz DF, Altura M, Hamilton MB, Weiss MR, et al. (2005) Molecular evolution of the chalcone synthase gene family and identification of the expressed copy in flower petal tissue of Viola cornuta. Plant Sci 168: 1127-1134.

48. Reddy AR, Scheffler B, Madhuri G, Srivastava MN, Kumar A, et al. (1996) Chalcone synthase in rice (Oryza sativa L.): Detection of the CHS protein in seedlings and molecular mapping of the chs locus. Plant Mol Biol 91:301-312.

49. Nakatsuka A, Izumi Y, Yamagishi M (2003) Spatial and temporal expression of chalcone synthase and dihydroflavonol 4-reductase genes in the Asiatic hybrid lily. Plant Sci 165: 759-767.

50. Jiang C, Schommer CK, Kim SY, Suh DY (2006) Cloning and characterization of chalcone synthase from the moss, Physcomitrella patens. Phytochemistry 67: 2531-2540.

51. Samappito S, Page J, Schmidt J, De-Eknamkul W, Kutchan TM (2002) Molecular characterization of root-specific chalcone synthases from Cassia alata. Planta 216: 64-71.

52. Wilmouth RC, Turnbull JJ, Welford RWD, Clifton IJ, Prescott AG, et al. (2002) Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana. Structure 10: 93-103.

53. Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat Struct Biol 6: 775-784.

54. Jez JM, Ferrer JL, Bowman ME, Austin MB, Schröder J, et al. (2001) Structure and mechanism of chalcone synthase-like polyketide synthases. J Ind Microbiol Biotechnol 27: 393-398.

55. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. (2008) Phylogeny. fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36: W465-469.

56. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381: 272.

57. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292: 195-202.

58. Tornielli (2009) The Genetics of Flower Color, petunia, 2nd ed. New York: Springer Science.

59. Vandesompele J, De Preter K, Poppe B, Van Roy N, De Paepe A (2002) Accurate normalization of real-time quantitative RT -PCR data by geometric averaging of multiple internal control genes 112.

60. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. (2009) The MIQE Guidelines: Minimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments Summary. Clin Chem 55: 611-622.

61. Shao H, Huang X, Zhang Y, Zhang C (2013) Main alkaloids of Peganum harmala L. and their different effects on dicot and monocot crops. Molecules 18: 2623-2634.

62. Martin CR (1993) Structure, function, and regulation of the chalcone synthase. Int Rev Cytol 147: 233-284.

63. Honda C, Kotoda N, Wada M, Kondo S, Kobayashi S, et al. (2002) Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. 40: 955-962.

64. Bombarely A, Mueller L Solanum lycopersicum Anthocyanidin synthase. , p.Accession Number: SOLYC11G013110.1 (LycoCyc).