Comprehensive influences of overexpression of a MYB transcriptor regulating anthocyanin biosynthesis on transcriptome and metabolome of tobacco leaves

Yuan Zong
Northwest Institute of Plateau Biology Chinese Academy of Sciences
Shiming Li
Northwest Institute of Plateau Biology Chinese Academy of Science
Xingyuan Xi
Northwest Institute of Plateau Biology Chinese Academy of Sciences
Dong Cao
Northwest Institute of Plateau Biology Chinese Academy of Sciences
Zhong Wang
National tobacco Research Center
Ran Wang
Natioal tobacco Research Center
Baolong Liu (blliu@nwipb.cas.cn)
Northwest Institute of Plateau Biology Chinese Academy of Sciences

Research article

Keywords: MYB transcriptor, secondary metabolism, anthocyanin, transcriptome, metabolome

Posted Date: July 6th, 2019

DOI: https://doi.org/10.21203/rs.2.11045/v1

Version of Record: A version of this preprint was published at International Journal of Molecular Sciences on October 16th, 2019. See the published version at https://doi.org/10.3390/ijms20205123.

Abstract

Background Overexpression of MYB transcription factors can induce the expression of structural genes for anthocyanin biosynthesis and increase the anthocyanin content of plant tissues. However, it remains unclear whether MYB transcription factor overexpression effects the activation of other genes and the concomitant accumulation of chemical compounds. Results Overexpression of LrAN2 promoted anthocyanin accumulation in a variety of tissues in tobacco cultivar Samsun. Only 185 unigenes, from total of 160,965, were expressed differently in leaves and 241 chemical compounds exhibited differences in accumulation. Four anthocyanins, including apigeninidin chloride, cyanidin 3-O-malonylhexoside, pelargonidin 3-O-beta-D-glucoside, and cyanidin 3,5-O-diglucoside were detected only in transgenic lines, which could explain the purple leaf phenotype. Beside anthocyanins, the phenylpropanoids, polyphenols (catechins), flavonoids, flavones, and flavonols were also upregulated. Overexpression of LrAN2 activated the basic helix-loop-helix transcription factor AN1b, and the MYB transcription factor MYB3. Additionally, structural genes associated with the phenylpropanoid biosynthetic pathway were activated, which lead to the upregulated accumulation of phenylpropanoid, polyphenol (catechin), flavonoid, flavone, flavonol, and anthocyanin. The MYB transcription factor CPC, a negative regulator of anthocyanin biosynthesis, was also expressed at increased levels in transgenic lines, which implie that a negative regulation mechanism existed in the anthocyanin biosynthesis pathway. The relative contents of all 19 differently accumulated amino groups and derivatives were decreased in transgenic lines, which meant that the phenylalanine biosynthesis pathway used other amino acids as substrates. Interestingly, the expression of acetylalkylglycerol acetylhydrolase was suppressed in transgenic lines, which caused the accumulation of 19 lyso-phosphatidylcholine derivatives and a decrease in production of eight octodecane derivatives. Conclusions Overexpression of LrAN2 activates the pathway of anthocyanin synthesis and metabolism in tobacco. Four anthocyanins lead to the purple leaf phenotype The main pathways of flavonoid biosynthesis were up-regulated. This research provides more information about the function of MYB transcription factors in anthocyanin biosynthesis and the production of other chemical compounds. This work will help breeders to obtain new plant cultivars with high anthocyanin contents using biotechnology.

Abstract

Background: Overexpression of MYB transcription factors can induce the expression of structural genes for anthocyanin biosynthesis and increase the anthocyanin content of plant tissues. However, it remains unclear whether MYB transcription factor overexpression effects the activation of other genes and the concomitant accumulation of chemical compounds.

Results: Overexpression of *LrAN2* promoted anthocyanin accumulation in a variety of tissues in tobacco cultivar Samsun. Only 185 unigenes, from total of 160,965, were expressed differently in leaves and 241 chemical compounds exhibited differences in accumulation. Four anthocyanins, including apigeninidin chloride, cyanidin 3-O-malonylhexoside, pelargonidin 3-O-beta-D-glucoside, and cyanidin 3,5-O-diglucoside were detected only in transgenic lines, which could explain the purple leaf phenotype. Beside

anthocyanins, the phenylpropanoids, polyphenols (catechins), flavonoids, flavones, and flavonols were also upregulated. Overexpression of *LrAN2* activated the basic helix-loop-helix transcription factor *AN1b*, and the MYB transcription factor *MYB3*. Additionally, structural genes associated with the phenylpropanoid biosynthetic pathway were activated, which lead to the upregulated accumulation of phenylpropanoid, polyphenol (catechin), flavonoid, flavone, flavonol, and anthocyanin. The MYB transcription factor *CPC*, a negative regulator of anthocyanin biosynthesis, was also expressed at increased levels in transgenic lines, which implie that a negative regulation mechanism existed in the anthocyanin biosynthesis pathway. The relative contents of all 19 differently accumulated amino groups and derivatives were decreased in transgenic lines, which meant that the phenylalanine biosynthesis pathway used other amino acids as substrates. Interestingly, the expression of acetylalkylglycerol acetylhydrolase was suppressed in transgenic lines, which caused the accumulation of 19 lyso-phosphatidylcholine derivatives and a decrease in production of eight octodecane derivatives.

Conclusions: Overexpression of *LrAN2* activates the pathway of anthocyanin synthesis and metabolism in tobacco. Four anthocyanins lead to the purple leaf phenotype The main pathways of flavonoid biosynthesis were up-regulated. This research provides more information about the function of MYB transcription factors in anthocyanin biosynthesis and the production of other chemical compounds. This work will help breeders to obtain new plant cultivars with high anthocyanin contents using biotechnology.

Keywords: MYB transcriptor, secondary metabolism, anthocyanin, transcriptome, metabolome

Background

Anthocyanins are glycosylated polyphenolic compounds that produce a range of colors varying from orange, red, and purple to blue in flowers, seeds, fruits, and vegetative tissues [1]. As anthocyanins are water-soluble pigments located mainly in cell vacuoles, their hue is influenced by the intravacuolar environment [2]. Over 600 anthocyanins have been identified in nature. The most common anthocyanins are the derivatives of six widespread anthocyanidins, namely, pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin. Anthocyanins can protect plants against various biotic and abiotic stresses [3], partially due to their powerful antioxidant properties. In addition, anthocyanin-rich food products have become increasingly popular, because of their attractive colors and suggested benefits to human health [4].

Anthocyanins emerge from a diverse family of aromatic molecules called flavonoids. Flavonoids comprise five major subgroups in higher plants beside anthocyanins: chalcones, flavones, flavonols, flavandiols or proanthocyanidins, and aurones (Winkel-Shirley, 2001a). Most of the structural genes encoding the enzymes responsible for anthocyanin biosynthesis have been isolated from various plants, for example, Arabidopsis (*Arabidopsis thaliana*), apple (*Malus domestica*), petunia (*Petunia hybrida*), grape (*Vitis vinifera*), and other species [5-8]. Conversion of the initial precursor phenylalanine is catalyzed in a stepwise manner by phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (*C4H*), 4-coumarate-CoA ligase (*4CL*), chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavanone 3-

hydroxylase (*F3H*), flavonoid 3-hydroxylase (*F3'H*) or flavonoid 3',5'-hydroxylase (*F3'5'H*), dihydroflavonol 4-reductase (*DFR*), anthocyanidin synthase (*ANS*), and O-methyltransferase (*OMT*) [9]. The transcription of the structural genes is regulated directly by activation of the MYB-bHLH-WD40 complex (MBW) that consists of R2R3-MYB, bHLH, and WD40 proteins [10]. Purple leaves or fruits rich in anthocyanin usually arise from the transcriptional activation of MYB transcription factors in plants [11]. In apples, red fruit color is controlled by *MdMYB10* [12], while the purple phenotype of sweet potato is strongly associated with *IbMYB1* [13]. *LhMYB6* and *LhMYB12* increase the expression levels of anthocyanin biosynthetic structural genes and determine the accumulation of anthocyanin in petunia (Petunia hybrida) [14].

Genetic engineering with only MYB transcription factors has been used to enhance anthocyanin accumulation in several plant species. Pattanaik reported that overexpression of MYB (*NtAn2*) in tobacco enhanced anthocyanin accumulation and expression levels of *CHS, CHI, F3H, DFR,* and *ANS* genes [15]; whereas, the expression of the MYB transcription factor *PtrMYB119* in hybrid poplar also enhanced the expression levels of anthocyanin biosynthetic genes and anthocyanin accumulation in a number of tissues [16]. Although it has been proven that the overexpression of MYB transcription factors promotes anthocyanin accumulation in plants, the molecular mechanism has not been clarified. The plant cell is a complex system. The overexpression of a transcription factor will activate structural genes and the expression of structural genes will produce some kind of chemical compound [17]. The plant cell will adapt to changes in the accumulation of chemical compounds within it. Previous researchers have reported on the transcription of a limited number of genes, in response to overexpression of MYB transcription factors, and the accumulation of a limited number of chemical compounds. These reports have not clarified the number of genes affected by the overexpression of MYB transcription factors and the turbulence effects of MYB transcription factor overexpression on chemical compounds other than anthocyanin.

Recently, the stepwise multiple ion monitoring-enhanced product ions (stepwise MIM-EPI) strategy was developed to analyze widely targeted metabolomes. The method can quantify hundreds of metabolites simultaneously in rice leaf [18]. Moreover, high-throughput RNA sequencing (RNA-Seq) has emerged as a powerful and cost-efficient tool for transcriptome analysis and transcript profiling in various plant species [19-21]. RNA-Seq has the advantage of providing comprehensive information on the nucleotide sequence of all genes expressed in the transcriptome. In this manuscript, these two technologies were employed to uncover the transcript and metabolite differences in the leaves of wild-type (WT) and transgenic lines with high anthocyanin content. Only 185 unigenes were expressed differently in the two lines, compared with 160,965 assembled unigenes, while 241 chemical compounds showed different levels of accumulation in the transgenic line and WT.

Results

Phenotype and anthocyanin content of transgenic line overexpressing *LrAN2*

As mentioned previously, *LrAN2* is a MYB transcription factor expressed only in the black fruit of *L. ruthenicum* [22]. Overexpression of *LrAN2* causes visible anthocyanin accumulation various parts of the plant (Fig. 1A). The relative anthocyanin contents of roots, stems, leaves, flowers, and seeds were significantly higher in the transgenic lines than the WT (Fig. 1B). Although anthocyanin differences are not visible in roots, the contents of this compound were found to be different in the two lines using chemical measurements. Because the anthocyanin content of the leaves was the highest, the leaves were chosen for further transcriptome and metabolome analysis.

Chemical compounds difference in transgenic lines and WT

The stepwise MIM-EPI strategy was developed recently to analyze widely targeted metabolomes. The overlapping display analysis of total ion flow charts (TIC) for MS detection and analysis of different quality control samples showed that the repeatability and reliability of the data were good enough to enable further analysis (Fig. S1). A total of 693 chemical compounds from 22 classes were detected and 224 compounds from 19 classes exhibited significantly different levels of accumulation in the transgenic lines and WT (Table 1). Transgenic lines displayed higher contents of 109 chemical compounds and the WT showed increased levels of 115 compounds (Table 1). Although 13 alcohols, two quinones, and five sterides were found, none of these accumulated differently in transgenic lines and the WT (Table S2).

The average ratio of differently-accumulated/total compounds was 35.05% (Table 1). The ratios of the nine classes were bigger than 35.05% (Table 1). The first seven classes included anthocyanins, polyphenols, flavonols, flavonoids, flavones, flavanones, and phenylpropanoids (Table 1). All these classes of chemical compounds belong to the flavonoid biosynthesis pathway. The phenylpropanoids, polyphenols, flavones, flavonols, flavonoids, and anthocyanins showed more upregulation in the transgenic lines, except for the compound flavone (Table 1). A total of nine anthocyanins with seven skeletons, peonidin, cyanidin, rosinidin, delphinidin, malvidin, pelargonidin, and apigeninidin, were identified (Table S2). Seven anthocyanins showed higher levels of accumulation in transgenic lines, while only one accumulated more highly in the WT (Table 1; Table S2). The raised accumulation of anthocyanin may be the reason for purple colored leaves in the transgenic lines. Four anthocyanins, including apigeninidin chloride, pelargonidin 3-O-beta-D-glucoside, callistephin chloride, and cyanidin 3,5-O-diglucoside (cyanin) were detected only in the transgenic lines (Table S2). The remaining three classes included lipids, vitamins and derivatives, and terpene. The class of compounds containing vitamins and their derivatives showed the same levels of up and downregulation (Table 1). Among the lipids, 11 compounds were more highly accumulated in transgenic lines, while 23 were reduced (Table S2). Interestingly, 19 compounds showing higher levels of production belonged to the lyso-phosphatidylcholines (and derivatives), while eight downregulated compounds belonged to the octodecanes (and derivatives; Table S2). Interestingly, all 16 differently accumulated amino acids exhibited lower levels of production in the transgenic lines (Table 1, Table S2).

Differently-expressed genes in transgenic lines and WT

Using transcriptome sequencing, the total number of bases from all samples was over 6 Gb after filtering (Table S3). The clean reads were further assembled into 160,965 unigenes. The average length of the unigenes was 1233 bp and the length of N50 was 2002 bp (Table S4). A Blast X search resulted in a total of 135,911 predicted proteins were predicted (Table S5).

Although a total 160,965 unigenes were found, only 185 were expressed differently in the transgenic lines and WT (Fig. 2A; Table S6). Compared with the WT, 39 unigenes were downregulated and 146 were upregulated (Fig. 2A; Table S6). Sixteen pathways contained only upregulated unigenes in the enriched 30 KEGG pathways (Fig. 2B). The pathways included cholesterol metabolism, cyanoamino acid metabolism, flavonoid biosynthesis, flavone and flavonol biosynthesis, galactose metabolism, glucosinolate biosynthesis, glutathione metabolism, the MAPK signaling pathway, nitrogen metabolism, phenylpropanoid biosynthesis, porphyrin and chlorophyll metabolism, starch and sucrose metabolism, and the sulfur relay system (Fig. 2B). Six pathways contained only downregulated unigenes (Fig. 2B). These pathways were lipid metabolism, fatty acid biosynthesis, fatty acid metabolism, monobactam biosynthesis, phagosome assembly, and SNARE interactions in vesicular transport (Fig. 2B).

The top thirty upregulated unigenes were homologous to the genes related to anthocyanin biosynthesis (Table S6), including *ANS*, *DFR, GST, LrAN2, difF, AN1b, ANP* (anthocyanin permease), *MYB3*, and so on. Apart from being related directly to anthocyanin biosynthesis, ANP is a multidrug resistance-associated protein that plays an important role in the transport of anthocyanin pigments into vacuoles [23], while difF encodes a cytochrome b5, which is required for full activity of flavonoid 3’5’-hydroxylase [24]. The log2FoldChange value of the largest upregulated unigene *ANS* reached 9.26 and this was followed by *DFR* (9.00; Table S6). The total FPKM of the unigenes homologous to *PAL, C4H, 4CL, CHS, CHI, F3H, F3'H, F3'5'H, DFR*, and *ANS* were compared in transgenic lines and the WT. These structural anthocyanin biosynthesis genes were significantly up-regulated, except for *4CL, F3H*, and *F3'5'H* (Fig. 3A; Table S7). The qPCR experiment confirmed the expression differences between these structural genes (Fig. 3B). Undoubtedly, the anthocyanin biosynthesis pathway is activated by the overexpression of MYB transcription factor *LrAN2*. Moreover, 10 upregulated unigenes were homologous to bHLH transcription factor *AN1b*, which has be proven to regulate anthocyanin biosynthesis [25](Table S6). Interestingly, ten and one unigenes homologous to transcription factors *MYB3* and *CPC*, respectively, showed more expression in transgenic lines (Table S6).

Conjoint analysis of transcriptome and metabolome

To futher evaluate the effects of the transcriptome change on the metabolome, the differently-expressed genes and differently-accumulated compounds were put on the KEGG pathway. Fourteen pathways contained both differently-expressed genes and differently-accumulated compounds (Fig. 4, Table S8). These were divided further into three classes: the phenylpropanoid second metabolism pathway, amino acid biosynthesis and metabolism pathway, and ether lipid metabolism pathway. The phenylpropanoid second biosynthesis pathway consisted of flavonoid biosynthesis, phenylpropanoid biosynthesis, and flavone and flavonol biosynthesis, while the amino acid biosynthesis and metabolism pathway consisted

of monobactam biosynthesis, the sulfur relay system, 2-oxocarboxylic acid metabolism, amino sugar and nucleotide sugar metabolism, glucosinolate biosynthesis, phenylalanine metabolism, cysteine and methionine metabolism, glycine, serine and threonine metabolism, aminoacyl-tRNA biosynthesis, and biosynthesis of amino acids. Only flavonoid biosynthesis could be enriched based on the p-value of the transcriptome and metabolome (Fig. 4). In the ether lipid metabolism pathway, acetylalkylglycerol acetylhydrolases showed lower expression in the transgenic lines than WT (Table S8). The acetylalkylglycerol acetylhydrolases may catalyze 1-alkyl-2-acetyl-sn-glycero-3-phosphocholine and H2O to produce 1-alkyl-sn-glycero-3-phosphocholine and acetate [26]. 1-alkyl-2-acetyl-sn-glycero-3-phosphocholine is the precursor of lyso-phosphatidylcholine and derivatives, while 1-alkyl-sn-glycero-3-phosphocholine is the precursor of octodecane and its derivatives [26]. The downregulation of the platelet-activating factor acetylhydrolase could explain why 19 lyso-phosphatidylcholines and derivatives were upregulated in transgenic lines and eight octodecanes and derivatives were produced at lower levels in metabolome analysis.

Discussion

In this manuscript, described the effect of *LrAN2* overexpression on the transcriptome and metabolome of tobacco leaves using stepwise MIM-EPI and RNA-seq. These were proven to be effective strategies for analyzing the comprehensive metabolome and transcriptome, respectively. A total of 693 chemical compounds from 22 classes was detected and 160,965 unigenes were assembled during transcriptome analysis.

Transcriptome turbulence was relatively simple and metabolome influence was extensive.

Interestingly, very few genes showed different levels of expression in transgenic lines and WT using transcriptome analysis. Although many unigenes were assembled, only 185 were shown to be expressed differently, which is a very low number. The top 30 unigenes showing upregulation were related to anthocyanin biosynthesis. These unigenes contained direct structural genes such as *ANS*, *DFR*, and transcription factors related to anthocyanin biosynthesis, but also ANP and diff, which are accessories that have not been identified in many plants. It could be speculated that overexpression of some regulation factors with RNA-seq analysis could help in the isolation of structural genes and regulation factors related to specific chemical compound biosynthetic pathways. In this case, a lot of differently-expression unigenes were annotated, particularly upregulated unigenes, but the functions of the remaining unigenes are still unknown. Genome editing technology was used to knock out these differently expressed genes to determine their functions.

Of 693 chemical compounds, the contents of 224 were significantly different in the transgenic lines and WT. The ratio of differently-accumulated compounds to total identified compounds reached 35.05%. Compared to that of the transcriptome, the influence of the metabolome was extensive.

Overexpression of MYB transcription factors influences phenylpropanoid metabolism.

In many plants, it has been noted that overexpression of a MYB transcription factor enhances the expression of anthocyanin biosynthetic genes as well as anthocyanin accumulation. This research further confirmed that the chemical classes of phenylpropanoids, polyphenols, flavones, flavonols, flavonoids, and flavanones showed high levels of accumulation (except anthocyanins) in transgenic lines. Moreover, overexpression of *LrAN2* could promote the bHLH transcriptor *AN1b*, which is related to anthocyanin biosynthesis. Transcriptional activation of the MBW complex may regulate the transcription of anthocyanin biosynthetic genes directly. *LrAN2* also promotes the transcription of other transcription factors, homologous to *MYB3* and *CPC*. Transcription factor *MYB3* is believed to regulate proanthocyanin (epiafzelechin and derivatives) biosynthesis [27], while *CPC* negatively regulates anthocyanin biosynthesis in Arabidopsis as it competes with the R2R3-MYB transcription factor *PAP1/2* [28] The accumulation of epiafzelechin and derivatives in the transgenic lines could be explained by the up-expression of *MYB3*. The upregulation of *CPC* due to anthocyanin accumulation is a mechanism for reducing anthocyanin biosynthesis.

Amino acid biosynthesis is influenced by the overexpression of MYB transcription factor.

Phenylalanine is the substrate for the phenylpropanoid biosynthesis pathway and it is an amino acid. Undoubtedly, phenylalanine completes for same substrates as other amino acids. The relative contents of all 19 differently accumulated amino acids and derivatives were decreased in the transgenic lines. In conjoint analysis of the transcriptome and metabolome, apart from the phenylpropanoid second biosynthesis pathway and fatty acid degradation pathway, the enriched KEGG pathways exhibiting different gene expression and chemical compound levels were related to amino acid biosynthesis and metabolism. The cyanoamino acid metabolism pathway derived from phenylalanine was active in the transcriptome. Although a chemical difference has not been found, this metabolic pathway is connected directly to those of glycine, serine and threonine, cystine and methionine, glutathione, and other amino acids.

Conclusion

This research describes a relatively comprehensive change in the metabolome and transcriptome of tobacco leaves following overexpression of a MYB transcription factor. Overexpression of *LrAN2* activates secondary metabolic pathways in tobacco, including transcription factors and compounds, and also affects the metabolism of amino acids and their derivatives. This study provides more information about the function of MYB transcription factors in anthocyanin biosynthesis and that of other chemical compounds. These results will help researchers to establish new plant cultivars containing high levels of anthocyanin using biotechnology.

Methods

Materials

The Nicotiana tabacum cultivar Samsun (voucher number: QTPMB-00031934) was chosen as a transformation plant. Seed was provided by John Innes Center and all plant materials were preserved in the Northwest Plateau Institute of Biology, Chinese Academy of Sciences. Chemical and molecular regents were purchased from Gansu Pengcheng Biological Technology Development Company.

Obtaining transgenic lines with overexpression of MYB transcription factor

LrAN2 is a MYB transcription factor isolated by our laboratory. This transcription factor is expressed only in the black fruit of Lycium ruthenicum and the allele variation of *LrAN2* is strictly related to the black fruit trait [22]. In this study, *LrAN2* was transformed into tobacco Samsun to obtain transgenic lines with high anthocyanin content. The vector PJAM1502 with a double 35s promoter was used for the transformation. The construct PJAM1502:LrAN2 was established using a Gateway Cloning Kit (Invitrogen, USA). Binary vectors were then transformed into Agrobacterium tumefaciens strain GV3101 using freeze-melt methods. The leaf disc transformation method was used for the tobacco transformation [29]. Regeneration tissues were grown on selective media containing 3% (w/v) sucrose, 0.7% (w/v) agar, 1.0 mg/L 6-benzylaminopurine (6-BA), 1.0 mg/L 1-naphthaleneacetic acid (NAA), and 150 mg/L kanamycin. The positive shoots were transferred to the greenhouse under long-day lighting (16 h light/8h dark). Screening of positive transgenic plants with specific primer Attb adapter for further experiments, the T3 family lines carrying the objective gene without separation were used.

Anthocyanin concentration

The “Total Monomeric Anthocyanin Pigment Content of Fruit Juices, Beverages, Natural Colorants, and Wines” (AOAC Official Method 2005.02) method was used to measure relative anthocyanin content in three independent experiments. HCL (1% v/v) was added to 100 mg of comminuted plant tissues (root, steam, leaf, seed, and flower), and the mixtures were incubated at 4°C overnight in the dark to extract anthocyanin. Statistical analyses were performed using the software package SPSS for Windows 11.5 with a 95% confidence interval [30].

Analysis of chemical contents

Transgenic lines and WT were cultured in a plant incubator for 1 month. The leaves were stripped and stored in a refrigerator at -80°C prior to chemical and transcriptome analyses. Freeze-dried leaves were crushed using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. 100 mg of the resulting powder was weighed and extracted overnight at 4°C with 1.0 ml 70% aqueous methanol. Following centrifugation at 10,000 g for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250 mg, 3 ml; ANPEL, Shanghai, China, www.anpel.com.cn/cnw) and filtered (SCAA-104, 0.22 μm pore size; ANPEL, Shanghai, China, http://www.anpel.com.cn/) before analysis by liquid chromatography-mass spectrometry (LCMS).

The sample extracts were analyzed using an LC-electrospray ionization-tandem MS (ESI MS/MS) system (high-performance LC [HPLC], Shim-pack UFLC SHIMADZU CBM30A system, www.shimadzu.com.cn/;

MS, Applied Biosystems 6500 Q TRAP, www.appliedbiosystems.com.cn/). The analytical conditions were as follows, HPLC: column, Waters ACQUITY UPLC HSS T3 C18 (1.8 μm, 2.1 mm*100 mm); solvent system, water (0.04% acetic acid): acetonitrile (0.04% acetic acid); gradient program, 95:5 v/v at 0 min, 5:95 v/v at 11.0 min, 5:95 v/v at 12.0 min, 95:5 v/v at 12.1 min, 95:5 v/v at 15.0 min; flow rate, 0.40 ml/min; temperature, 40°C; injection volume: 2 μL. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS.

Linear ion trap (LIT) and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP, API 6500 Q TRAP LC/MS/MS system) equipped with an ESI turbo ion-spray interface, operating in a positive ion mode and controlled by Analyst 1.6.3 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 500°C; ion spray voltage (IS) 5500 V; ion source gas I (GSI), gas II (GSII), and curtain gas (CUR) were set to 55, 60, and 25.0 psi, respectively; the collision gas (CAD) was high. Instrument tuning and mass calibration were performed with 10 and 100 μmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were acquired as multiple reaction monitoring (MRM) experiments with the collision gas (nitrogen) set to 5 psi. Declustering potential (DP) and collision energy (CE) for individual MRM transitions was done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period. Statistical analyses were performed using the software package SPSS for Windows 11.5 with a 95% confidence interval [30].

Illumina sequencing and data analysis

Total RNA was extracted using an RNAprep Pure Plant Kit (Tiangen Company, Beijing, China). The quality of the total RNA was checked by electrophoresis and the concentration was determined by NanoDrop (Thermo Scientific, Wilmington, DE, USA). The cDNA libraries were created according to the preparation method for mRNA-seq samples (Illumina Inc, San Diego, CA, USA).

Using an Illumina HiSeq 2000 instrument (Illumina Inc, San Diego, CA, USA), the DNA library was sequenced by Novogene with three repeats. Before sequence assembly, the original reading was filtered to remove ambiguous, joint and low quality sequences, so as to establish high quality readings. After purity filtration, de novo assembly of the transcriptome into unigenes was carried out using Trinity, a short-read assembly program [31]. The unigene sequences were compared to protein databases using Blast X (e-value < 0.00001) in the following order: Nr, Swiss-Prot, KEGG, and COG. The blast results were used to extract the CDS from unigene sequences and to translate them into peptide sequences. The blast results were also used to train ESTScan [32]. The expression levels of the unigenes were calculated from the FPKM (fragments per kb per million reads) value. The difference in unigenes between purple and green leaf transcripts was analyzed by IDEG6 software [33]. The false discovery rate (FDR) method was introduced to determine the threshold p-value at FDR $\leq$ 0.001, | log2 ratio > 1 was determined the significance of the differential expression of unigenes. All differentially abundant unigenes were mapped to the GO and KEGG pathway databases and then the respective numbers of unigenes for every GO and KEGG orthology (KO) term were calculated. To compare these unigenes with the whole transcriptome

background, significantly enriched GO and KO terms from the set of differentially abundant unigenes were identified using the hypergeometric test [34].

Quantitative reverse-transcription-PCR

The cDNA was synthesized from the same total RNA for Illumina sequencing using a reverse transcription kit (Thermo Fisher First Strand cDNA Synthesis Kit, Beijing, China). The β-ACTIN gene was selected for cDNA normalization and several differently-expressed genes were selected to confirm the results of RNA-seq. The primers for the selected genes were designed by Primer 5.0 (Table S1). The qPCR was conducted with the Premix Ex Taq™ probe (TaKaRa, China) in Applied Biosystems QuantStudio3 (Thermo Fisher Company, Beijing, China). The thermal cycle for qPCR was 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, and 60°C for 34 s; the last stage was 95°C for 15 s, 60°C for 1 min, and 95°C for 15 s. Each plate was repeated three times. The 2-ΔΔCT method was used to analyze the expression levels of the different genes [35].

Abbreviations

PAL: phenylalanine ammonia lyase, C4H: cinnamate 4-hydroxylase, 4CL: 4-coumarate-CoA ligase, CHS: chalcone synthase, CHI: chalcone isomerase, F3H: flavanone 3-hydroxylase, F3'H: flavonoid 3-hydroxylase, F3'5'H: flavonoid 3',5'-hydroxylase, DFR: dihydroflavonol 4-reductase, ANS: anthocyanidin synthase, OMT: O-methyltransferase, MBW: MYB-bHLH-WD40 complex, RNA-Seq: RNA sequencing, WT: wild-type, TIC: total ion flow charts, 6-BA: 6-benzylaminopurine, NAA: 1-Naphthaleneacetic acid, HPLC: high-performance LC, LIT: Linear ion trap, QQQ: quadrupole, IS: ion spray voltage, GSI: ion source gas I, GSII: gas II, CUR: curtain gas, CAD: collision gas, DP: Declustering potential, CE: collision energy.

Declarations

Acknowledgements

Not applicable

Author contributions

Y.Z., R.W. and B.L.L. conceived and designed the study. Y.Z., S.M.L., X.Y.X., D.C and Z.W. performed the experiments. S.M.L., X.Y.X. D.C and Z.W. analyzed data. B.L.L. wrote the paper. All authors read and approved the final manuscript.

Funding

The sequencing was supported by the Project of Qinghai Science and Technology Department (2018-NK-133, 2019-ZJ-982Q), the editing and publishing fee were supported by West Light Talent Program of the Chinese Academy of Sciences and The Innovative Academy of Seed Design, Chinese Academy of Sciences.

Availability of data and materials

All data generated or analysed during this study are included within the article and its additional files.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Additional file

Additional file1: Fig. S1. Detection of TIC overlaps by QC sample Mass Spectrometry. (DOCX 87 Kb)

Additional file2: Table S1. The primers used in this manuscript. (DOCX 15 Kb)

Additional file3: Table S2. The differently-accumulation chemical compounds in leaves of transgenic lines and WT. (XLSX 40 Kb)

Additional file4: Table S3. Summary of transcriptome sequencing data. (DOCX 15 Kb)

Additional file5: Table S4. Distribution of the lengths of the assembled unigenes. (DOCX 16 Kb)

Additional file6: Table S5. The annotation of assembled unigenes. (DOCX 15 Kb)

Additional file7: Table S6. The differently-expression unigenes in leaves of transgenic lines and WT. (XLSX 57 Kb)

Additional file8: Table S7. The relative transcript level of the unigenes relative to anthocyanin biosynthesis. (DOCX 19 Kb)

Additional file9: Table S8. The KEGG pathways of differently expression unigenes and differently accumulation chemical compounds. (DOCX 22 Kb)

References

1. Asen S, Stewart RN, Norris KH: Copigmentation of Anthocyanins in Plant-Tissues and Its Effect on Color. *Phytochemistry* 1972, 11(3):1139-&.

2. Florenciano EG, Calderon AA, Munoz R, Barcelo AR: Patterns of Anthocyanin Deposition in Vacuoles of Suspension Cultured Grapevine Cells. *Phyton-Int J Exp Bot* 1992, 53(1):47-50.

3. Ahmed I, Cooper K, McGrath R, Griego G, Poole MS, Hanisch RJ: A Process Study of the Development of Virtual Research Environments. *Astr Soc P* 2014, 485:325-328.

4. Pojer E, Mattivi F, Johnson D, Stockley CS: The Case for Anthocyanin Consumption to Promote Human Health: A Review. *Compr Rev Food Sci F* 2013, 12(5):483-508.

5. Li SS, Strid A: Anthocyanin accumulation and changes in CHS and PR-5 gene expression in Arabidopsis thaliana after removal of the inflorescence stem (decapitation). *Plant Physiol Bioch* 2005, 43(6):521-525.

6. Honda C, Kotoda N, Wada M, Kondo S, Kobayashi S, Soejima J, Zhang ZL, Tsuda T, Moriguchi T: Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. *Plant Physiol Bioch* 2002, 40(11):955-962.

7. Sun W, Meng XY, Liang LJ, Li YQ, Zhou TT, Cai XQ, Wang L, Gao X: Overexpression of a Freesia hybrida flavonoid 3-O-glycosyltransferase gene, Fh3GT1, enhances transcription of key anthocyanin genes and accumulation of anthocyanin and flavonol in transgenic petunia (Petunia hybrida). *In Vitro Cell Dev-Pl* 2017, 53(5):478-488.

8. Zhu ZG, Li GR, Liu L, Zhang QT, Han Z, Chen XS, Li B: A R2R3-MYB Transcription Factor, VvMYBC2L2, Functions as a Transcriptional Repressor of Anthocyanin Biosynthesis in Grapevine (Vitis vinifera L.). *Molecules* 2019, 24(1).

9. Li YD, Meng FL, Zheng YN: Study of Anthocyanins in fruit of different Vaccinium genotypes. *Proceedings of the 8th International Symposium on Vaccinium Culture* 2006(715):589-593.

10. Lloyd A, Brockman A, Aguirre L, Campbell A, Bean A, Cantero A, Gonzalez A: Advances in the MYB-bHLH-WD Repeat (MBW) Pigment Regulatory Model: Addition of a WRKY Factor and Co-option of an Anthocyanin MYB for Betalain Regulation. *Plant and Cell Physiology* 2017, 58(9):1431-1441.

11. Chiu LW, Zhou X, Burke S, Wu X, Prior RL, Li L: The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiol* 2010, 154(3):1470-1480.

12. Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC: Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J* 2007, 49(3):414-427.

13. Kim CY, Ahn YO, Kim SH, Kim YH, Lee HS, Catanach AS, Jacobs JM, Conner AJ, Kwak SS: The sweet potato IbMYB1 gene as a potential visible marker for sweet potato intragenic vector system. *Physiol Plant*

2010, 139(3):229-240.

14. Yamagishi M, Shimoyamada Y, Nakatsuka T, Masuda K: Two R2R3-MYB genes, homologs of Petunia AN2, regulate anthocyanin biosyntheses in flower Tepals, tepal spots and leaves of asiatic hybrid lily. *Plant Cell Physiol* 2010, 51(3):463-474.

15. Pattanaik S, Kong Q, Zaitlin D, Werkman JR, Xie CH, Patra B, Yuan L: Isolation and functional characterization of a floral tissue-specific R2R3 MYB regulator from tobacco. *Planta* 2010, 231(5):1061-1076.

16. Cho JS, Nguyen VP, Jeon HW, Kim MH, Eom SH, Lim YJ, Kim WC, Park EJ, Choi YI, Ko JH: Overexpression of PtrMYB119, a R2R3-MYB transcription factor from Populus trichocarpa, promotes anthocyanin production in hybrid poplar. *Tree Physiology* 2016, 36(9):1162.

17. Li Y, Tu K, Zheng S, Wang J, Li Y, Hao P, Li X: Association of feature gene expression with structural fingerprints of chemical compounds. *J Bioinform Comput Biol* 2011, 9(4):503-519.

18. Wei C, Liang G, Guo Z, Wang W, Zhang H, Liu X, Yu S, Xiong L, Jie L: A Novel Integrated Method for Large-Scale Detection, Identification, and Quantification of Widely Targeted Metabolites: Application in the Study of Rice Metabolomics. *Mol Plant* 2013, 6(6):1769-1780.

19. Zhang N, Liu B, Ma C, Zhang G, Chang J, Si H, Wang D: Transcriptome characterization and sequencing-based identification of drought-responsive genes in potato. *Mol Biol Rep* 2014, 41(1):505-517.

20. Dorn KM, Fankhauser JD, Wyse DL, Marks MD: De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *The Plant Journal* 2013, 75(6):1028-1038.

21. Firon N, LaBonte D, Villordon A, Kfir Y, Solis J, Lapis E, Perlman TS, Doron-Faigenboim A, Hetzroni A, Althan L: Transcriptional profiling of sweetpotato (Ipomoea batatas) roots indicates down-regulation of lignin biosynthesis and up-regulation of starch biosynthesis at an early stage of storage root formation. *BMC Genomics* 2013, 14(1):460.

22. Zong Y, Zhu X, Liu Z, Xi X, Li G, Cao D, Wei L, Li J, Liu B: Functional MYB transcription factor encoding gene AN2 is associated with anthocyanin biosynthesis in Lycium ruthenicum Murray. *BMC Plant Biol* 2019, 19(1):169.

23. Christopher Dean G, Paula C, Virginia W: A multidrug resistance-associated protein involved in anthocyanin transport in Zea mays. *Plant Cell* 2004, 16(7):1812-1826.

24. Vetten N, De, Horst J, Ter, Schaik HP, Van, Boer A, De, Mol J, ., Koes R, . A cytochrome b5 is required for full activity of flavonoid 3', 5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower

colors. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96(2):778-783.

25. Vincenzo DA, Riccardo A, Giorgia B, Immacolata C, Mar CM, Ana Beatriz CS, Pasquale C, Carlo F, Francesca P, Domenico C: High AN1 variability and interaction with basic helix-loop-helix co-factors related to anthocyanin biosynthesis in potato leaves. *Plant Journal* 2015, 80(3):527-540.

26. Blank ML, Lee T, ., Fitzgerald V, ., Snyder F, . A specific acetylhydrolase for 1-alkyl-2-acetyl-sn-glycero-3-phosphocholine (a hypotensive and platelet-activating lipid). *Journal of Biological Chemistry* 1981, 256(1):175-178.

27. Zhu Q, Sui S, Lei X, Yang Z, Lu K, Liu G, Liu YG, Li M: Ectopic Expression of the Coleus R2R3 MYB-Type Proanthocyanidin Regulator Gene SsMYB3 Alters the Flower Color in Transgenic Tobacco. *PLoS One* 2015, 10(10):e0139392.

28. Hui-Fen Z, Karen F, Abha K, Kranz RG: CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in Arabidopsis. *Mol Plant* 2009, 2(4):790-802.

29. Horsch RB, Fry JE, Hoffmann NL, Eichholtz D, Rogers SG, Fraley RT: A Simple and General Method for Transferring Genes into Plants. *Science* 1985, 227(4691):1229-1231.

30. Lee J, Durst RW, Wrolstad RE: Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method: collaborative study. *J AOAC Int* 2005, 88(5):1269-1278.

31. Cabau C, Escudie F, Djari A, Guiguen Y, Bobe J, Klopp C: Compacting and correcting Trinity and Oases RNA - Seq de novo assemblies. *Peerj* 2017, 5.

32. Iseli C, Jongeneel CV, Bucher P: ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In*: 1999*. 138-148.

33. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA: IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol Genomics* 2003, 12(2):159-162.

34. Rivals I, Personnaz L, Taing L, Potier M-C: Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* 2007, 23(4):401-407.

35. Hurley JH, Dasgupta A, Andrews P, Crowell AM, Ringelberg C, Loros JJ, Dunlap JC: A Tool Set for the Genome-Wide Analysis of Neurospora crassa by RT-PCR. *G3 (Bethesda)* 2015, 5(10):2043-2049.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS7.pdf
- Fig.2.jpg
- TableS8.pdf
- Fig.1.jpg
- Table1..pdf
- Fig.3.jpg
- TableS1.Theprimersusedinthismanuscript.pdf
- TableS4.pdf
- TableS5.Theannotationofassembledunigenes.pdf
- TableS3.Summaryoftranscriptomesequencingdata.pdf
- TableS2.xlsx
- TableS6.xlsx
- Fig.4.jpg
- Fig.S1.TheTICoverlapsbyQCsampleMassSpectrometry.pdf