Higher anthocyanin accumulation was associated with higher transcription levels of anthocyanin biosynthesis genes in spinach
Higher anthocyanin accumulation was associated with higher transcription levels of anthocyanin biosynthesis genes in spinach

Xiaofeng Cai, Lihao Lin, Xiaoli Wang, Chenxi Xu, Quanhua Wang*

Development and Collaborative Innovation Center of Plant Germplasm Resources, Shanghai Engineering Research Center of Plant Germplasm Resources, College of Life and Environment Science, Shanghai Normal University, Shanghai, 200234, China

*Corresponding authors:
Quanhua Wang, Email: wangquanhua@shnu.edu.cn; Telephone and fax number: +86 21 64321102;

Development and Collaborative Innovation Center of Plant Germplasm Resources, Shanghai Engineering Research Center of Plant Germplasm Resources, College of Life and Environment Science, Shanghai Normal University, Shanghai, 200234, China
Abstract

Spinach (*Spinacia oleracea* L.) is widely cultivated as an economically important green leafy vegetable crop for fresh and processing consumption. The red purple spinach shows abundant anthocyanin accumulation in the leaf and leaf petiole. However, the molecular mechanisms of anthocyanin synthesis in this species are still undetermined. In the present study, we investigated the pigments formation and identified anthocyanin biosynthetic genes in spinach, and performed the expression analysis of anthocyanin related genes in the purple and green cultivar by quantitative PCR. Results showed that accumulation of anthocyanin was the dominant pigment resulting in the red coloration in spinach, and 22 biosynthesis genes and 25 regulatory genes were identified in spinach, based on the spinach genomic and transcriptomic database. Furthermore, the expression patterns of genes encoding enzymes indicated that *SoPAL*, *SoUFGT3* and *SoUFGT4* were possible candidate genes for anthocyanin biosynthesis in red purple spinach. The expression patterns of transcription factors indicated that two *SoMYBs*, three *SobHLHs* and one *SoWD40* were drastically up-regulated and co-expression in red purple spinach, suggesting an essential role of regulatory genes in the anthocyanin biosynthesis of spinach. The results above enhanced our understanding about the molecular mechanisms of anthocyanin biosynthesis in purple spinach.

Keywords:

Spinach; anthocyanin biosynthesis; structural genes; transcription factors; *Spinacia oleracea* L;
Abbreviations

ANS, Anthocyanin synthase; bHLH, Basic helix loop helix; CHI, Chalcone isomerase; CHS, Chalcone synthase; DAFB, Days after full bloom; DFR, DihydroXavonol 4-reductase; DEPC, Diethylpyrocarbonate; F3H, Flavanone 3-hydroxylase; PAL, Phenylalanine ammonialyase; SE, Standard error; TF, Transcript factor; CT, Threshold cycle; UFGT, UDP-glucose: flavonoid-3-O-glucosyltransferase; MBW, MYB-bHLH-WD40 protein complex; qPCR, Quantitative reverse transcription polymerase chain reaction.

Introduction

Color is one of the most important determinants of vegetable crop quality. The red coloration is an attractive feature in many vegetables, fruits, and other plant tissues, and is associated with anthocyanin accumulation. Anthocyanins, as an important subclass of flavonoids, are the main water-soluble pigments, playing an important role in physiological processes in plants; Anthocyanins are also powerful antioxidants and could be induced under stress conditions or infection by pathogens (Zhang et al. 2014). Furthermore, More and more evidences indicate that dietary foods consumptions with rich anthocyanins are beneficial to human health and are associated with protection against certain cancers, cardiovascular diseases and other chronic human disorders (de Pascual-Teresa et al. 2010; Li et al. 2017).

It is well-known that anthocyanins are synthesized through a branch of the phenylpropanoid pathway, and the structural genes that directly participate in the process of anthocyanin accumulation have been well studied in many species (Shi and Xie 2014; Zhang et al. 2014). Anthocyanins are formed by series genes including beginning genes, early biosynthetic genes
(EBGs) and late biosynthetic genes (LBGs) (Shi and Xie 2014). The beginning genes include three genes, phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumaryl CoA ligase (4CL), which are the general phenylpropanoid pathway; The early biosynthetic genes (EBGs) include chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H) and flavanone 3’-hydroxylase (F3’H). The late biosynthesis genes (LBGs) include DFR (dihydrofavonol 4-reductase), anthocyanidin synthase/leucoanthocyanidin dioxygenase (ANS/LDOX), and UDP-glucose:flavonoid 3-Oglucosyl transferase (UFGT), which are core pathway for anthocyanin biosynthesis; besides, UFGT transferring of glucoside to the 3-O position of flavonoids, is a key enzyme in the regulation of anthocyanin biosynthesis in the fruit of many crops. After biosynthesis, anthocyanins are transported to vacuoles or cell walls, to exhibit their brilliant colors (Jaakola 2013; Shi and Xie 2014; Zhang et al. 2014).

The structural genes involved in core anthocyanin biosynthesis are known to be coordinately regulated by a specific transcriptional complex, known as MBW (MYB-bHLH-WD40) protein complex, which forms by R2R3 MYB transcription factors, MYC-like basic helix-loop-helix (bHLH) and WD40-repeat proteins, and binds to promoters and activates transcription of structural genes of the late anthocyanin biosynthetic pathway (Jaakola 2013; Xu et al. 2015b; Zhang et al. 2014). The role of MBW in anthocyanin biosynthesis has been elucidated in plant (Rahim et al. 2014; Zhang et al. 2014). In Arabidopsis, R2R3-MYB genes PAP1, PAP2, MYB113, and MYB114, bHLH family genes TT8, GL3, EGL3, and WD40 family gene TTG1 are recognized as the key genes encoding respective components of MBW complex. While early biosynthesis genes are mainly controlled by R2R3-MYB and bHLH individually or coordinate with others proteins (Shi and Xie 2014; Zhang et al. 2014). Besides, it has also been reported that others
proteins or transcription factors could also activate anthocyanin biosynthesis in plants, such as CBP60g, NAC, CCoAOMT, GSTs (Mahmood et al. 2016; Perez-Diaz et al. 2016; Shaipullah et al. 2016; Zhou et al. 2015; Zou et al. 2017).

Spinach (*Spinacia oleracea* L.) is widely cultivated as an economically important green leafy vegetable crop for fresh and processing consumption (van Treuren et al. 2011). Spinach is produced in more than 50 countries, primarily in China. The annual worldwide gross production of spinach was approximately 23 million tonnes in 2013, of which around 91% was produced in China (FAOSTAT, http://faostat3.fao.org). Spinach is considered one of the healthiest vegetables in the human diet due to its high concentration of nutrients and health-promoting compounds (Correll et al. 2011; Lester et al. 2013). Recently, few red-purple leaf spinach are introduced and cultivated in China. Compared to green cultivars of spinach, the red cultivars, which display abundant anthocyanin accumulation in the stems and leaves, could attract more attention from consumers due to their brilliant color and high levels of health-promoting ingredients. However, the molecular mechanisms of anthocyanin accumulation in this species are still poorly understood. Because of the previous absence of genome information, little is known about the genes involved in the anthocyanin biosynthetic pathways in spinach. Recently, the spinach genome and transcriptome has been sequenced and de novo assembled, which provide an important tool for identifying and classifying structural and regulatory genes involved in anthocyanin biosynthesis (Dohm et al. 2014; Xu et al. 2017; Xu et al. 2015a). To analyze spinach leaves color formation at the molecular level, we systemically investigated the pigments formation in spinach and analyzed the transcriptomic changes between purple and green leaves. The results above enhanced our understanding about the molecular mechanisms of anthocyanin biosynthesis in purple spinach.
Materials and methods

Materials

Seeds of green spinach variety (SP2) were provided by Laizhou Seed Company (Shandong, China) and seeds of red spinach accession (SP106) were obtained from the North Central Regional Plant Introduction Station, Ames, Iowa, respectively. Both cultivated spinach plants were grown under standard green-house conditions with a 16-hour light (27°C) and 8-hour dark (15°C) cycle. The fresh healthy full expanding leaves and petioles were randomly collected from six seedlings of 60-day-old plants, and immediately frozen in liquid nitrogen and stored at -80°C till use.

Determination of pigments content

Anthocyanin content of samples was determined using the protocol of Teng et al. (2005). Briefly, frozen, homogenized samples (0.2 g) were extracted for 1 d at 4 °C in 5 mL of 1% (v/v) hydrochloric acid in methanol. The mixture was centrifuged at 10,000 rpm for 15 min at 4 °C and the absorbance of the supernatant was measured at 530 and 657 nm. Relative anthocyanin concentrations were calculated with the formula modification: 

$$Q_{\text{Anthocyanins}} = (A_{530} - 0.25 \times A_{657}) \times M^{-1} \times MW \times 1000 \times E^{-1},$$

where $Q_{\text{Anthocyanins}}$ is the amount of anthocyanins, $A_{530}$ and $A_{657}$ are the absorption values at the indicated wavelengths, $M$ fresh weight (g) of the tissues, $MW$ is the molecular weight of cyanidin-3-glucoside in 449.2; and $E$ is the molar absorptivity, which equal to 26, 900 for cyanidin-3-glucoside. The relative anthocyanin amount was defined as the product of relative anthocyanin concentration and extraction solution volume (Teng et al. 2005). The chlorophyll and carotenoids content calculations were followed as described by Cai et al. (2015) (Cai et al. 2015). Total phenolics and polyphenol contents were carried out using the method described by Lewis et al (Lewis et al. 1998). The obtained data were
the mean of three independent replicates.

**Identification of anthocyanin biosynthesis and regulatory genes**

All the sequences of anthocyanin biosynthesis and regulatory genes were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/). The spinach genome and transcriptome sequences (http://www.spinachbase.org/cgi-bin/spinach/index.cgi) were used to identify the anthocyanin biosynthesis and regulation genes in spinach using BLASTN and BLASTP with a cut off E-value ≤ 1E−10 and coverage ≥ 0.75. All of the putative genes obtained were submitted to Pfam to confirm the presence of the relevant domain. The theoretical isoelectric point (pI) and molecular weight (Mw) were identified by ProtParam tool (http://web.expasy.org/protparam/).

**Sequence Alignment and Phylogenetic Analysis**

Multiple alignments of protein sequences from spinach were performed using the ClustalW program. A phylogenetic tree was constructed by the neighbor-joining method using MEGA4 (Tamura et al. 2007), with the Poisson correction, random seed of phylogeny test, and pairwise deletion option parameters enabled. The reliability of the trees obtained was tested using a bootstrap analysis with 1,000 replicates; only clades with a test value higher than 50 were selected for the consensus tree. Images of the phylogenetic trees were also drawn using MEGA4.

**RNA extraction and real-time qRT-PCR analysis**

Total RNA was extracted with TRIzol reagent (Invitrogen, USA) and treated with RNase-free DNase I (Invitrogen, Gaithersburg, MD, USA). First strand cDNA was synthesized from 3 mg total RNA from each sample using a high capacity cDNA reverse transcription kit (TOYOBO, Japan), according to the supplier's protocols. Quantitative RT-PCR (qRT-PCR) was then conducted as described previously (Cai et al. 2013). The spinach So18s gene was used as the internal control.
Amount of the transcripts in the green cultivar SP2 was set at 1.0. A total of 34 gene specific primer pairs for anthocyanin biosynthesis and regulation genes were designed using Primer Premier 5, and were synthesised from Sangon Biotech (Shanghai) Co., Ltd. The primers used are listed in Suppl.

**Statistical analysis**

Statistical analysis was performed using Excel and SAS software. Significant differences were calculated using the Student’s $t$ test at 95% confidence limit.

**Results**

**Accumulation of anthocyanins in different spinach species**

The concentrations of anthocyanins, chlorophylls, and carotenoids in leaves and leaf petioles of two different spinach cultivars were measured, and the total phenolics and polyphenol contents were detected. Results showed that the anthocyanins contents of leaves and leaf petioles of red purple variety SP106 were 37.1 and 40.7 µg/g FW, respectively. Comparatively, the anthocyanins contents of green leaf variety SP2 were almost undetected (Fig 1A). Besides, significant difference of total phenolics and polyphenol contents was also found between the leaf of SP106 and SP2 (Fig 1B). Whereas, there is no different of the contents of chlorophyll a, chlorophyll b and carotenoid between two spinach varieties (Supplb).

**Identification of anthocyanin biosynthesis genes in spinach**

All the genes involved in the anthocyanin biosynthesis were identified, and 22 biosynthesis genes were classified and identified in spinach, based on the spinach genomic and transcriptomic database. The names and the unigene number of identified genes, the length of ORF and amino
acid, the genomic location, introns numbers, molecular weight and isoelectric point are shown in Table 1.

There are five SoUFGT genes, four SoCHS genes, two genes each of So4CL, SoF3H, SoFLS and SoANS, and one gene each of SoPAL, SoC4H, SoCHI, SoF3'H and SoDFR (Table 1). Of the 22 genes, half of them were mapped to the six chromosomes of spinach, with four and three genes located on chromosomes chr6 and chr1, respectively, and one gene located on remaining chromosomes (Table 1). The remaining 11 genes were anchored on 10 different scaffolds, which have not yet been mapped onto a chromosome.

The predicted protein sequence of SoUFGTs was aligned with others UFGTs involved in anthocyanin regulation in other species, and the phylogenetic tree was constructed (Fig. 2). Sequence alignment analysis suggested that the UFGTs proteins all have a glycosyl transferase domain, which is a necessary motif for UFGTs (Fig. 2A). Phylogenetic analysis of UFGTs showed that UFGTs were formed three clearly separates clades (Fig. 2B). SoUFGT1 and SoUFGT2 showed high similarity to AtUGT79B1 (56.9 % and 56.1 % amino acid identity, respectively), and clustered into one clade. SoUFGT3 shared 62.9 % identity with StUFGT, SoUFGT4 and SoUFGT5 shared 46.8 % and 34.8 % identity with PaUFGT, respectively (Fig. 2).

**Identification and analysis of anthocyanin regulatory genes in spinach**

A total of 76 spinach R2R3-MYBs and 125 Arabidopsis R2R3-MYBs were used to construct a phylogenetic tree (Supplc). SoMYBs were integrated with AtMYBs in clustered phylogenetic clades or subclades and divided into 26 MYB subgroups overall (Suppld). Accordingly, there was three SoMYBs in Sg4 (Spo18626, Spo17369 and Spo13499, renamed as SoMYB3, SoMYB4 and SoMYB5, respectively), one SoMYB in Sg5 (Spo25305, renamed as SoMYB6) and two SoMYBs
in Sg7 (Spo17636 and Spo01215, renamed as SoMYB1 and SoMYB2, respectively) (Fig. 3A),
and their deduced proteins both have MYB-like domains (Fig. 3B); no SoMYBs were found in
Sg6 and Sg15. The predicted protein sequence was aligned with MYBs known to control
anthocyanin gene expression in other species (Fig. 3). Among the Sg4 members, SoMYB3,
SoMYB4 and SoMYB5 clustered together with PtMYB057, AmMYB330 and PhMYB4, sharing
with 55.9 %, 56.4 % and 62.5 % identity, respectively, and shared high identity with GmMYBZ2
and GmMYB48. Sg7 members SoMYB1 and SoMYB2 clustered together and showed closest
homology to VvMYBF1 and DkMYB4, respectively. Sg5 member SoMYB6 clustered together
with AT5G35550 and OsMYB3 (Fig. 3).

Phylogenetic analysis was performed using 107 spinach SobHLHs and 151 AtbHLHs, which
were divided into 32 subfamilies (Suppld). Consequently, 15 SobHLHs involved in regulation of
flavonoid or anthocyanin metabolism were identified, including six SobHLHs in subfamily2
(Spo01207, Spo01158, Spo02407, Spo02408, Spo23296 and Spo26018, renamed as
SobHLH1-SobHLH6, respectively), one SobHLH in subfamily5 (Spo04675, renamed as
SobHLH7) and eight SobHLHs in subfamily24 (Spo01327, Spo08670, Spo08635, Spo21887,
Spo10465, Spo20776, Spo14554 and Spo08210, renamed as SobHLH8-SobHLH15, respectively)
(Fig. 4A), and their deduced proteins all have a helix-loop-helix domain, which is a necessary
motif for bHLHs (Fig. 4B). Phylogenetic analysis suggested that each of SobHLHs fell in the
clade with the orthologous of bHLHs known to control anthocyanin gene expression in other
species other species (Fig. 4).

A total of 159 WD genes in spinach and the 237 WDRs in Arabidopsis were used to construct
a phylogenetic tree (Supple). The phylogenetic trees analysis showed that there was two
SoWD40s clustered together with TRANSPARENT TESTA GLABRA 1 (TTG1) (Walker et al. 1999) (Spo02820 and Spo13615, renamed as SoWD40-1 and SoWD40-2, respectively), and two SoWD40s clustered together with CONSTITUTIVELY PHOTOMORPHOGENIC1 (COP1) (Lau and Deng 2012) (Spo22375 and Spo12020, renamed as SoWD40-3 and SoWD40-4, respectively), involving in the anthocyanin biosynthesis pathway. Phylogenetic analysis manifested that SoWD40-1 and SoWD40-2 clustered together with AtCOP1 and MdCOP1, and showed 78.7 % and 61.4 % identity with AtCOP1; 67.0 % and 60.5 % identity with MdCOP1, respectively. SoWD40-3 and SoWD40-4 showed high homology with PpTTG1 and GhTTG2, sharing with 78.6 % and 88.2 % identity, respectively (Fig. 5 and Supplf).

In total, 25 possible regulatory genes (6 SoMYBs, 15 SobHLHs and 4 SoWD40s TFs) were identified as candidates to be the regulators of the anthocyanin accumulation in spinach, and the detailed informations are shown in Supplg.

Expression patterns of anthocyanin biosynthesis genes in different cultivars

In order to investigate the mechanisms underlying the anthocyanin accumulation in red purple spinach, the transcripts of anthocyanin biosynthetic enzymes and regulatory genes were examined in the leaves and leaf petioles of the two cultivars by qPCR. Except for SoDFR and SoUFGT1, the expression of anthocyanin biosynthetic genes SoPAL, SoC4H, So4CL1, SoCHSs, SoCHI, SoF3Hs, SoF3'H, SoFLSs, SoANSs, and SoUFGTs are shown in Figure 6. Consistent with pigment production in leaves, the expression of all the anthocyanin structural genes was drastically upregulated in SP106 comparison to SP2; besides, the transcript levels of genes encoding enzymes SoPAL, So4CL1 and SoUFGTs were much higher than those of other biosynthetic genes. Among those upregulated genes, So4CL1 exhibited the highest fold increase (500 folds at least) in the
leaves of SP106; SoPAL, So4CL2, SoUGT3 and SoUGT4 displayed high expression levels (200 folds changes at least) in the leaves of SP106; the expression levels of SoUGT2, SoUGT5, SoCHI, SoF3'H, SoFLS1, SoC4H, SoCHS1 and SoCHS2 showed at least 15 fold changes between SP106 and SP2 (Fig. 6). Whereas in leaf petioles, the relative expression level of SoPAL, SoUGT2, SoUGT3 and SoUGT4 were significantly upregulated in SP106, comparing to SP2. The others structural genes showed no significant difference among two cultivars or even down-regulated expressing in SP106 (Fig. 6).

**Expression profiles of the transcription factors in different cultivars**

To verify whether any of the regulatory genes controlling the transcription of anthocyanin structural genes in spinach, the transcripts of identified MYBs, bHLHs, and WD40s were examined. In the results shown in Figure 7, the expression of SoMYB1 and SoMYB5 was detectable in all tested samples being highest in leaves of SP106, with more than eight fold higher than that of SP2, and displayed slightly changes between the petioles of SP106 and SP2 (Fig. 7). The transcript level of SoMYB4 in leaves was more than 15 fold higher than that in leaves of SP2, but was remarkable lower in petioles of SP106 than that in SP2 (Fig. 7). The expression of SoMYB2, SoMYB3 and SoMYB6 was no detectable in two spinach cultivars.

The four SoWD40s were found to have a higher expression profile in leaves of SP106 than that in SP2. The expression levels of SoWD40-1 and SoWD40-4 in leaves of SP106 were about 108- and 8-fold higher than those in SP2, respectively, and showed no significant difference between two cultivars. While the expression levels of SoWD40-2 and SoWD40-3 in leaves of SP106 were about 6- and 321-fold higher than those in SP2, respectively, but was remarkable lower in petioles of SP106 than that in SP2 (Fig. 7).
The relative expression level of most SobHLHs was detectable in all tested samples, except for six SobHLHs genes (SobHLH3, SobHLH4, SobHLH6, SobHLH7, SobHLH11 and SobHLH12). The expression of SobHLHs presented similar expression patterns with SoWD40s and SoMYBs, displaying high expression levels in leaves of SP106, and showing at least 20-fold higher than those in SP2. The expression levels of SobHLH5, SobHLH9, SobHLH10, SobHLH13, SobHLH14 and SobHLH15 in petioles of SP106 were extraordinary higher than in SP2. There were no significant differences in the transcript level of SobHLH1 and SobHLH8 among petioles of two cultivars.

Discussion

Spinach is cultivated worldwide as an economically important green leafy vegetable crop for fresh and processing consumption (van Treuren et al. 2011), and is considered one of the healthiest vegetables in the human diet due to its high concentration of nutrients and health-promoting compounds (Correll et al. 2011; Lester et al. 2013). Anthocyanins are an important quality for vegetables, and research into anthocyanin biosynthesis may facilitate improvement of the quality of vegetables productions. However, the anthocyanin accumulations and molecular mechanisms of anthocyanin synthesis are not well studied. In the study, we investigated the pigments formation in two spinach cultivars, identified candidate genes, and detected the expression changes of genes involved in anthocyanin accumulations in two tissues. The results above enhanced our understanding about the molecular mechanisms of anthocyanin biosynthesis in spinach.

Anthocyanins are responsible for red coloration in spinach

Anthocyanins, chlorophylls, and carotenoids are the most important pigments in the color
formation of plants. In this study, we measured the contents of anthocyanin, chlorophyll, carotenoids, total phenolics and polyphenol in leaves and leaf petioles. The results showed that red purple cultivar showed remarkable differences of anthocyanins, total phenolics and polyphenol contents comparing to green cultivars (Fig 1). However, the contents of chlorophyll a, chlorophyll b and carotenoid are no different between two spinach varieties (Supplb). Those results indicated that the accumulation of anthocyanins is the main pigment responsible for red coloration in spinach.Previous reports have showed that the high amount of anthocyanins in the purple cultivars indicates a high ability to synthesize and accumulate anthocyanins, and the drastic differences in anthocyanin accumulation arise from cultivar and genetic specificity (Rahim et al. 2014; Yang et al. 2015).

**Anthocyanins biosynthesis genes in spinach**

The completion of spinach genome sequencing and released transcriptome has provided researchers with an opportunity to identify genes involved in anthocyanin biosynthesis (Dohm et al. 2014; Xu et al. 2015a). In the present study, we identified 22 putative anthocyanin biosynthetic genes (Table 1) that encoded enzymes controlling anthocyanin biosynthesis. Five *SoUFGT* genes, four *SoCHS* genes, two genes each of *So4CL*, *SoF3H*, *SoFLS* and *SoANS* were identified in spinach genome (Table 1). Similar phenomena have been observed in other plants. UFGTs, CHSs, 4CLs, F3Hs, FLSs and ANSs each had several copies (Guo et al. 2014; Wu et al. 2014), even existed with small gene families, such as six FLSs members in Arabidopsis (Owens et al. 2008), and four UFGT genes in the apple (Meng et al. 2015), which may lead to functional difference at different tissues in different cultivars.

UFGT, responsible for the transfer of glucoside to the 3-O position of flavonoids, is a key
enzyme in the regulation of anthocyanin biosynthesis. Five SoUFGT genes were identified in the
spinach genome in this study, sequence alignment analysis suggested that the SoUFGT proteins
have a glycosyl transferase domain (Fig. 2A), and showed high similarity to UFGTs involved in
anthocyanin synthesis in other plants (Hu et al. 2011; Yonekura-Sakakibara et al. 2012). Phylogenetic analysis also revealed close relationships among UFGTs of anthocyanin accumulation plants (Fig. 2B), suggesting that SoUFGTs may be the key enzyme genes controlling the biosynthesis of anthocyanin in spinach.

In plants, anthocyanin accumulation is regulated at the transcriptional level by the MBW transcriptional activation complex. The most MYB, bHLH and WD40 genes involved in anthocyanin biosynthesis cluster in the same phylogenetic clades or subgroups (Sg), such as members of Sg4, Sg5, Sg6, Sg7 and Sg15 (Hichri et al. 2011; Zhao et al. 2013); subfamilies 2, 5, and 24 were involved in regulation of flavonoid or anthocyanin metabolism (Carretero-Paulet et al. 2010). Therefore, six SoMYBs, 15 SobHLHs and two SoWD40s were identified (Supplementary Fig. 2-4). All the SoMYBs, SobHLHs and SoWD40s had conservation domains, and showed closest homology to known genes controlled anthocyanin gene expression (Fig. 3-5).

**Expression of enzyme genes involved in anthocyanin biosynthesis**

The anthocyanin biosynthetic pathway involves at least ten enzyme genes. We compared the expression of these genes in leaves and leaf petioles between two different cultivars. Consistent with pigment production, the expression levels of the early and late biosynthetic genes in leaves of two cultivars showed the same tendency with anthocyanin accumulations, especially the early biosynthetic genes, **SoPAL** and **So4CLI**, and late biosynthetic genes, **SoUFGT3** and **SoUFGT4**, displayed 200 to 500 folds changes (Fig. 6). Correlation of the expression of key biosynthetic
genes with anthocyanin and flavonoid accumulation is fully documented in many species (Shi and Xie 2014; Zhang et al. 2014). It was well established that anthocyanins are synthesized via the flavonoid pathway, which is a branch of the phenylpropanoid pathway, starts from the key amino acid phenylalanine to produce 4-coumaroyl CoA by PAL, and ends up with anthocyanidin to produce anthocyanin by UFGT (Jaakola 2013; Zhang et al. 2014). PAL is the first committed enzyme in the phenylpropanoid pathway, and high expression of PAL account for the accumulation of anthocyanins and pal1 pal2 double mutants exhibited a reduced production of anthocyanins and proanthocyanidins (Huang et al. 2010; Saito et al. 2013). On the other side, UFGT plays a major role in the stable accumulation of anthocyanins, the high expression of the anthocyanin biosynthetic gene UFGT and high UFGT activity correlated with the anthocyanin concentration in plant tissues (Meng et al. 2015; Yang et al. 2015; Zhai et al. 2016), and the mutant of UFGT was characterized as exhibiting drastically decreased anthocyanin contents (Saito et al. 2013). In conclusion, SoPAL, SoUFGT3 and SoUFGT4 are the possible candidate genes of anthocyanin biosynthesis in spinach.

**Expression of transcription factors involved in anthocyanin biosynthesis**

The core transcriptional complex of MYB, bHLH, and WD40 regulates the expression of the anthocyanin biosynthesis genes as verified in plant (Jaakola 2013; Saito et al. 2013; Zhang et al. 2014). To verify whether any of the regulatory genes controlling the transcription of anthocyanin structural genes, the transcripts of identified regulatory genes were examined. The results shown that three SoMYBs, nine SobHLHs and four SoWD40s were the possible regulatory genes greatly upregulated in leaves rich with anthocyanins, but displayed distant expression levels in leaf petioles (Fig. 7), suggesting an essential role of regulatory genes in the anthocyanin biosynthesis.
of spinach. In addition, based on the statistical analysis (data not shown), we found a positive correlation between the expression levels of the transcription factors SoMYB4 and SoMYB5 with SobHLH13, SobHLH14, SobHLH15 and SoWD40-1 in spinach. Furthermore, these results were consistent with results of other species, the transcripts of MYBs, bHLHs and WDs were positively correlated with anthocyanin accumulation in different tissues or stages, indicating different anthocyanin biosynthesis regulation mechanisms between tissues or development stages (Meng et al. 2015; Yang et al. 2015; Zhai et al. 2016). Thus, further research is needed to determine which regulatory genes could regulate anthocyanin accumulation.

Conclusion

We investigated the pigments formation and identified anthocyanin biosynthetic genes in spinach, and performed the expression analysis of anthocyanin related genes in the purple and green cultivar. The accumulation of anthocyanins is the main pigment responsible for red coloration in spinach. Besides, 47 possible anthocyanins biosynthesis genes were identified in spinach, and the enzyme genes SoPAL, SoUFGT3 and SoUFGT4 were possible candidate genes for anthocyanin biosynthesis in red purple spinach. In addition, two SoMYBs, three SobHLHs and one SoWD40 were drastically up-regulated and co-expression in red purple spinach, suggesting an essential role of regulatory genes in the anthocyanin biosynthesis of spinach. The results above enhanced our understanding about the molecular mechanisms of anthocyanin biosynthesis in purple spinach. The potential mechanism behind the regulation of anthocyanin biosynthesis is complex, and further studies of gene function involved still need to be explored.
Acknowledgements
This work was supported by National Natural Sciences Foundation of China (No. 31501754), Scientific Research Innovation Project of Shanghai Science and Technology Committee (16391901000), Shanghai Engineering Research Center of Plant Germplasm Resources (17DZ2252700) and Development and Collaborative Innovation Center of Shanghai (ZF1205).

Conflict of interest
The authors declare no conflicts of interest. All authors read and approved the final manuscript.

Compliance with ethical standards
This article does not contain any studies with human subjects or animals performed by any of the authors.

References


doi: DOI 10.1007/s10658y010y9713yy.


https://mc06.manuscriptcentral.com/genome-pubs
Figure legend:

Figure 1. The concentrations of anthocyanin, the total phenolics and polyphenol contents in leaves and leaf petioles of two spinach cultivars. (A) Total anthocyanin content analysis of leaves and leaf petioles of two spinach cultivars. (B) Total total phenolics and polyphenol analysis of leaves and leaf petioles of two spinach cultivars. Number 1-4 represent the leaf of SP2, leaf of SP106, leaf petioles of SP2 and leaf petioles of SP106, respectively. Error bars represent standard deviation among three independent replicates. Data are the mean ± SD of three independent replicates. The statistical significance of the differences between samples was calculated with ANOVA by paired-group comparisons. Different lowercase letters represent significant differences at $P < 0.05$.

Figure 2. Sequence alignment (A) and phylogenetic analysis (B) of SoUFGTs and other proteins involved in the regulation of the anthocyanin pathway from a range of species. The protein sequences were aligned using ClustalW method and the tree was constructed by neighbor-joining method and 1000 bootstrap replicates using MEGA version 5. Numbers close to the nodes indicate bootstrap values and scale shows 0.1 amino acid substitutions per site. The upper solid triangular squares represent for spinach identified anthocyanin regulating proteins. GenBank accession numbers were MdUFGT (KF711858.1), VvUFGT (AAB81683.1), BrUFGT (XP_009108098.1), BnUFGT (CDY27124.1), StUFGT (XP_006358760.1), SpUFGT (XP_015088519.1), SlUFGT (XP_004247894.1), PgUFGT (GU371443.1), BoUFGT (EF531098), PaUFGT (GQ325589), AtUGT79B1 (AT5G54060.1), AtUGT75C1 (AT4G14090.1) and AtUGT78D2 (AT5G17050.1).
Figure 3. Phylogenetic analysis (A) and sequence alignment (B) of SoMYBs and other MYBs involved in the regulation of the anthocyanin pathway from a range of species. The protein sequences were aligned using ClustalW method and the tree was constructed by neighbor-joining method and 1000 bootstrap replicates using MEGA version 5. Numbers close to the nodes indicate bootstrap values and scale shows 0.1 amino acid substitutions per site. The upper solid triangular squares represent for spinach identified anthocyanin regulating proteins. GenBank accession numbers were PtMYB14 (ABD60279), DvMYB2 (BAJ33514), HlMYB1 (CAI46244), PhMYB4 (ADX33331), EgMYB1 (CAE09058), GmMYBZ2 (ABI73970), GhMYB9 (AAS92347), AmMYB308 (P81393), AmMYB330 (P81395), GmMYB48 (ABH02823), PtMYB057 (EEF00115), ZmMYB42 (CAJ42204), PgMYB16 (ACN12958), GhMYB38 (AAK19618), OsMYB3 (BAA23339), ZmC1 (AAA33482), VvMYB5a (AAS68190), DkMYB4 (BAI49721), VvMYBF1 (ACV81697), GhMYB1 (CAD87007), LjMYB12 (BAF74782), MdMYB22 (AAZ20438), ZmP (P27898) and AtMYB5 (At3g13540).

Figure 4. Phylogenetic analysis (A) and sequence alignment (B) of SobHLHs and other bHLHs involved in the regulation of the anthocyanin pathway from a range of species. The protein sequences were aligned using ClustalW method and the tree was constructed by neighbor-joining method and 1000 bootstrap replicates using MEGA version 5. Numbers close to the nodes indicate bootstrap values and scale shows 0.1 amino acid substitutions per site. The upper solid triangular squares represent for spinach identified anthocyanin regulating proteins. GenBank accession numbers were HbMYC4 (AEG74014), Re_02518914 (XP_002518914), Pt_002299425 (XP_002299425), VvMYC4 (XP_002279973), GmMYC2-like (XP_003528771), AtbHLH3.
(O23487), AtbHLH14 (O23090), AtbHLH17 (Q9ZPY8), AtbHLH26 (Q9FE22), VvPIF3-like
(XP_002276198), Md_AEX32796 (AEX32796), VITISV_042277 (CAN83346), Pp_SPATULA-like (ADG56590), Re_002514702 (XP_002514702), CrMYC5 (ACM41588), AtbHLH24 (Q9FUA4), AtbHLH8 (O80536), AtbHLH16 (Q8GZ38), AtbHLH72 (Q570R7),
AtbHLH119 (Q8GT73), AtbHLH127 (Q7XHI7), AtbHLH23 (Q9SVU6), AtbHLH15 (Q8GZM7), AtbHLH9 (Q8W2F3), AtbHLH65 (Q8GZM7), AtbHLH124 (Q84LH8) and AtbHLH132 (Q8L5W7).

Figure 5. Phylogenetic analysis of SoWD40s and other WD40s involved in the regulation of the anthocyanin pathway from a range of species. The protein sequences were aligned using ClustalW method and the tree was constructed by neighbor-joining method and 1000 bootstrap replicates using MEGA version 5. Numbers close to the nodes indicate bootstrap values and scale shows 0.1 amino acid substitutions per site. The upper solid triangular squares represent for spinach identified anthocyanin regulating proteins. GenBank accession numbers were MdTTG1 (ADI58760), PFWD (BAB58883), ZmPAC1 (AAM76742), PhAN11 (AAC18914), InWDR (BAE94407), AtTTG1 (CAB45372), MdCOP1 (BAM08276.1), MIWD40a (AHJ80982.1), StAN11 (AEF01097.1), VvWD40 (XP_002282158), MtWD40-1 (ABW08112.1), OsWD40 (XP_466030), GhTTG2 (AAM95644.1) and PpTTG1 (ACQ65867).

Figure 6. Expression analysis of anthocyanin biosynthetic genes in two tissues of the two spinach cultivars. Number 1-4 represent the leaf of SP2, leaf of SP106, leaf petioles of SP2 and leaf petioles of SP106, respectively. Error bars represent standard deviation among three independent
replicates. Data are the mean ± SD of three independent replicates. The statistical significance of
the differences between samples was calculated with ANOVA by paired-group comparisons.
Different lowercase letters represent significant differences at $P < 0.05$.

**Figure 7.** Expression analysis of anthocyanin regulatory genes in two tissues of the two spinach
cultivars. Number 1-4 represent the leaf of SP2, leaf of SP106, leaf petioles of SP2 and leaf
petioles of SP106, respectively. Error bars represent standard deviation among three independent
replicates. Data are the mean ± SD of three independent replicates. The statistical significance of
the differences between samples was calculated with ANOVA by paired-group comparisons.
Different lowercase letters represent significant differences at $P < 0.05$. 
<table>
<thead>
<tr>
<th>Gene name</th>
<th>Unigene number</th>
<th>Chromosome/Scaffold and location</th>
<th>Strand</th>
<th>Introns</th>
<th>CDS (bp)</th>
<th>Size (aa)</th>
<th>Molecular weight(kD)</th>
<th>Isoelectric point (pI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SoPAL</td>
<td>Spo16743</td>
<td>Super_scaffold_153</td>
<td>Rev</td>
<td>1</td>
<td>2166</td>
<td>721</td>
<td>78.40</td>
<td>6.03</td>
</tr>
<tr>
<td>SoC4H</td>
<td>Spo20748</td>
<td>Super_scaffold_88</td>
<td>Rev</td>
<td>2</td>
<td>1515</td>
<td>504</td>
<td>57.73</td>
<td>9.26</td>
</tr>
<tr>
<td>So4CL1</td>
<td>Spo25427</td>
<td>chr2</td>
<td>For</td>
<td>4</td>
<td>1698</td>
<td>565</td>
<td>62.14</td>
<td>5.41</td>
</tr>
<tr>
<td>So4CL2</td>
<td>Spo10338</td>
<td>chr3</td>
<td>Rev</td>
<td>5</td>
<td>1722</td>
<td>573</td>
<td>62.15</td>
<td>5.64</td>
</tr>
<tr>
<td>SoCHS1</td>
<td>Spo26612</td>
<td>SpoScf_01236</td>
<td>Rev</td>
<td>1</td>
<td>1179</td>
<td>392</td>
<td>43.19</td>
<td>6.35</td>
</tr>
<tr>
<td>SoCHS2</td>
<td>Spo26611</td>
<td>SpoScf_01236</td>
<td>Rev</td>
<td>1</td>
<td>1176</td>
<td>391</td>
<td>42.98</td>
<td>6.15</td>
</tr>
<tr>
<td>SoCHS3</td>
<td>Spo27151</td>
<td>SpoScf_02197</td>
<td>For</td>
<td>1</td>
<td>1185</td>
<td>394</td>
<td>43.47</td>
<td>6.23</td>
</tr>
<tr>
<td>SoCHS4</td>
<td>Spo25740</td>
<td>chr1</td>
<td>For</td>
<td>1</td>
<td>1176</td>
<td>391</td>
<td>43.07</td>
<td>6.15</td>
</tr>
<tr>
<td>SoCHI</td>
<td>Spo18570</td>
<td>chr4</td>
<td>Rev</td>
<td>3</td>
<td>753</td>
<td>250</td>
<td>26.90</td>
<td>5.08</td>
</tr>
<tr>
<td>SoF3H1</td>
<td>Spo16254</td>
<td>SpoScf_01777</td>
<td>For</td>
<td>2</td>
<td>1110</td>
<td>369</td>
<td>41.35</td>
<td>5.33</td>
</tr>
<tr>
<td>SoF3H2</td>
<td>Spo26379</td>
<td>SpoScf_01201</td>
<td>For</td>
<td>2</td>
<td>1119</td>
<td>372</td>
<td>42.22</td>
<td>5.26</td>
</tr>
<tr>
<td>SoF3'H</td>
<td>Spo05258</td>
<td>chr5</td>
<td>For</td>
<td>2</td>
<td>1560</td>
<td>519</td>
<td>57.72</td>
<td>8.74</td>
</tr>
<tr>
<td>SoFLS1</td>
<td>Spo10830</td>
<td>SpoScf_00206</td>
<td>For</td>
<td>2</td>
<td>1008</td>
<td>335</td>
<td>38.06</td>
<td>5.77</td>
</tr>
<tr>
<td>SoFLS2</td>
<td>Spo22544</td>
<td>SpoScf_02761</td>
<td>For</td>
<td>2</td>
<td>858</td>
<td>285</td>
<td>32.21</td>
<td>5.56</td>
</tr>
<tr>
<td>SoDFR</td>
<td>Spo25888</td>
<td>chr6</td>
<td>Rev</td>
<td>5</td>
<td>1029</td>
<td>342</td>
<td>38.58</td>
<td>5.50</td>
</tr>
<tr>
<td>SoANS1</td>
<td>Spo26936</td>
<td>chr6</td>
<td>Rev</td>
<td>1</td>
<td>1059</td>
<td>352</td>
<td>39.86</td>
<td>5.38</td>
</tr>
<tr>
<td>SoANS2</td>
<td>Spo26937</td>
<td>chr6</td>
<td>Rev</td>
<td>0</td>
<td>510</td>
<td>169</td>
<td>19.29</td>
<td>4.73</td>
</tr>
<tr>
<td>SoUFGT1</td>
<td>Spo03893</td>
<td>chr1</td>
<td>For</td>
<td>0</td>
<td>1380</td>
<td>459</td>
<td>51.24</td>
<td>5.83</td>
</tr>
<tr>
<td>SoUFGT2</td>
<td>Spo25984</td>
<td>chr6</td>
<td>For</td>
<td>0</td>
<td>1347</td>
<td>448</td>
<td>49.41</td>
<td>5.79</td>
</tr>
<tr>
<td>SoUFGT3</td>
<td>Spo13657</td>
<td>SpoScf_03439</td>
<td>For</td>
<td>0</td>
<td>1422</td>
<td>473</td>
<td>53.62</td>
<td>4.89</td>
</tr>
<tr>
<td>SoUFGT4</td>
<td>Spo23078</td>
<td>Super_scaffold_143</td>
<td>For</td>
<td>1</td>
<td>1392</td>
<td>463</td>
<td>50.65</td>
<td>5.88</td>
</tr>
<tr>
<td>SoUFGT5</td>
<td>Spo15163</td>
<td>chr1</td>
<td>Rev</td>
<td>1</td>
<td>1335</td>
<td>444</td>
<td>49.70</td>
<td>6.43</td>
</tr>
</tbody>
</table>
Captions for Supplementary Material:

**gen-2017-0261SupplA:** Characteristics of the predicted anthocyanin regulatory genes in spinach.

**gen-2017-0261SupplB:** Sequences of primers used for qRT-PCR in this study.

**gen-2017-0261SupplC:** The contents of chlorophyll a, chlorophyll a and carotenoid in two spinach cultivars. Number 1-4 represent the leaf of SP2, leaf of SP106, leaf petioles of SP2 and leaf petioles of SP106, respectively. Error bars represent standard deviation among three independent replicates. Data are the mean ± SD of three independent replicates. Lowercase letters represent significant differences at $P < 0.05$.

**gen-2017-0261SupplD:** Phylogenetic tree showing relationships between Arabidopsis and spinach MYB TFs. The upper solid triangular squares represent for Arabidopsis anthocyanin regulating MYB TFs. Phylogenetic and molecular evolutionary analyses were conducted using neighbor-joining method and 1000 bootstrap replicates using MEGA version 6. Numbers close to the nodes indicates bootstrap values.

**gen-2017-0261SupplE:** Phylogenetic tree showing relationships between Arabidopsis and spinach bHLH TFs. The upper solid triangular squares represent for Arabidopsis anthocyanin regulating bHLH TFs. Phylogenetic and molecular evolutionary analyses were conducted using neighbor-joining method and 1000 bootstrap replicates using MEGA version 6. Numbers close to the nodes indicates bootstrap values.

**gen-2017-0261SupplF:** Phylogenetic tree showing relationships between Arabidopsis and spinach WD40 TFs. The upper solid triangular squares represent for Arabidopsis anthocyanin regulating WD40 TFs. Phylogenetic and molecular evolutionary analyses were conducted using neighbor-joining method and 1000 bootstrap replicates using MEGA version 6. Numbers close to the nodes indicates bootstrap values.
Alignment of WD40s from spinach and other species. Full-length amino acid sequences of WD40s in spinach and other species aligned by ClustalX. GenBank accession numbers were MdTTG1 (AD158760), PFWD (BAB58883), ZmPAC1 (AAM76742), PhAN11 (AAC18914), InWDR (BAE94407), AtTTG1 (CAB45372), MdCOP1 (BAM08276.1), MiWD40a (AHJ80982.1), StAN11 (AEF01097.1), VvWD40 (XP_002282158), MtWD40-1 (ABW08112.1), OsWD40 (XP_466030), GhTTG2 (AAM95644.1) and PpTTG1 (ACQ65867).