Isolation and Characterization of Key Genes that Promote Flavonoid Accumulation in Purple-leaf Tea (Camellia sinensis L.)

Xiujuan He1,2, Xuecheng Zhao1, Liping Gao1, Xingxing Shi1, Xinlong Dai2, Yajun Liu1, Tao Xia2 & Yunsheng Wang1,2

There were several high concentrations of flavonoid components in tea leaves that present health benefits. A novel purple-leaf tea variety, ‘Mooma1’, was obtained from the natural hybrid population of Longjing 43 variety. The buds and young leaves of ‘Mooma1’ were displayed in bright red. HPLC and LC-MS analysis showed that anthocyanins and O-Glycosylated flavonols were remarkably accumulated in the leaves of ‘Mooma1’, while the total amount of catechins in purple-leaf leaves was slightly decreased compared with the control. A R2R3-MYB transcription factor (CsMYB6A) and a novel UGT gene (CsUGT72AM1), that were highly expressed in purple leaf were isolated and identified by transcriptome sequencing. The over-expression of transgenic tobacco confirmed that CsMYB6A can activate the expression of flavonoid-related structural genes, especially CHS and 3GT, controlling the accumulation of anthocyanins in the leaf of transgenic tobacco. Enzymatic assays in vitro confirmed that CsUGT72AM1 has catalytic activity as a flavonol 3-O-glucosyltransferase, and displayed broad substrate specificity. The results were useful for further elucidating the molecular mechanisms of the flavonoid metabolic fluxes in the tea plant.

Flavonoids are produced from the phenylpropanoid and malony-CoA pathway, are a large group of plant secondary compounds, which fulfill important functions such as defending pathogen infection2, avoiding the damage from UV irradiation3, and interaction with the plant with temperature and other environments4. There are mainly six groups of flavonoids in plant tissues, including flavan-3-ols (catechins and proanthocyanidins), anthocyanins, flavanonols, flavonols, flavones and phenolic acid5. The key components of flavonoids are responsible for the attractive colors seen in various plant organs (such as leaf, flower, and fruit), avoiding excess light damage, and aiding pollination and seed dispersal6. In recent years, the health benefit of anthocyanins got more significant attention. Scientists have systematically investigated and confirmed that anthocyanins could reduce the risk of lifestyle-related diseases, such as hypertension6, liver disorder7, cerebral disorder8, dysentery and diarrhea9, and urinary problems10.

Tea (Camellia sinensis) is the most widely consumed non-alcoholic beverage. It contains rich catechins (flavan-3-ol) flavonoids, and around 12–24% of the dry mass of the younger leaves11. Nevertheless, the concentration of anthocyanins is few to zero in normal tea leaves. Recently, researchers have developed special purple-leaf tea varieties in different tea growing countries12-15. These special tea varieties have been shown to contain high quantity of anthocyanins. There are eight anthocyanins that are isolated and identified from anthocyanin-rich tea14. The health benefits of purple-leaf tea have been preliminarily studied16,17. Hsu et al. study confirmed that the anthocyanin-rich tea is considered as a novel dietary compound for colorectal cancer chemoprevention16. However, there was no major systematic or global analysis reported on the molecular mechanism of anthocyanin accumulation in these tea varieties.

Flavonoids are derived from the phenylpropanoid metabolic pathway, in which cinnamoyl-(C6-C3) is combined to yield the backbone with three malonyl-CoA units. In recent years, most of the functional and regulatory

1School of Life Science, Anhui Agricultural University, Hefei, Anhui, 230036, China. 2State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, Anhui, 230036, China. Xiujuan He, Xuecheng Zhao and Liping Gao contributed equally to this work. Correspondence and requests for materials should be addressed to T.X. (email: xiatao62@126.com) or Y.W. (email: wang_li09213@sina.com)
genes involved in these flavonoid pathways are evident through model species. In this pathway, the functional enzymes have been classified into two groups, early-biosynthetic enzymes are required for the synthesis of flavonoids, and late-biosynthetic enzymes are used for the synthesis of anthocyanins, catechins, and flavonols. UDP-glycosyltransferase (UGT) is the last enzyme produced during the anthocyanin and flavonol biosynthesis. It helps to catalyze the transfer of glucosyl moiety from UDP-glucose to produce the first stable pigment. However, because of the competition in the biosynthetic pathways for substrate and multiplex branches, the relationship among plant flavonoids still remain unclear.

Several studies have reported the polymorphic characteristics of some key genes, which in turn determined the fluxion of flavonoid pathway. Numerous studies have suggested that the fluxion of this pathway is tightly regulated by the transcription factor complex MYB-bHLH-WD40(MBW). MBW complexes are also known to control various aspects of epidermal cell patterning, such as the development of trichomes and root hairs. Specific combinations of R2R3-MYB transcription factor with bHLH and WD40 regulate specific pathways of anthocyanin or PA biosynthesis. The accumulation of anthocyanins requires the activity of MBW complex consisting of MYB (including PAP1, PAP2, MYB113, MYB114), TT8, and TTG1 in Arabidopsis. Some studies have indicated that R2R3-MYB transcription factors played a central role in distinguishing the target gene in these pathways. Zhao et al. had predicted that R2R3-MYB genes were involved in the flavonoid biosynthesis of C. sinensis. However, C. sinensis lacked Sg6 CsMYB genes, which might be due to low anthocyanin content in the tea plants.

According to the report from a latest study by Sun et al., isolation of transcription factor R2R3-MYB anthocyanin 1 (CSAN1) from purple-leaf tea variety ‘Zijuan’ conferred ectopic accumulation of anthocyanins in purple-leaf tea. In this study, a purple-leaf tea variety ‘Mooma 1’ was obtained from the natural hybrid population of Longjing 43 variety by asexual reproduction. ‘Mooma 1’ plants were mainly accumulated and abundant in the palisade mesophyll (Pa) and xylem (Xy) cells, which were devoid of upper epidermal cells.

**Comparison of leaf color and flavonoid concentrations of different tea varieties.** As shown in Fig. 1a, anthocyanins were notably accumulated in ‘Mooma 1’ leaves. The color of young leaves and stems of ‘Mooma 1’ was bright red, whereas the color of wild-type was yellowish green (Fig. 1c). The sections of tea leaves were sliced via free-hand sectioning to avoid the loss of anthocyanins. As shown in Fig. 1e and g, anthocyanins in the purple-leaf tea variety were mainly accumulated and abundant in the palisade mesophyll (Pa) and xylem (Xy) cells, which was different from the wild type.

**Comparison of flavonoid compositions of different tea varieties.** To investigate whether the flavonoid biosynthesis pathway of purple-leaf tea variety was different from the wild type, the flavonoid components of the fresh leaves were extracted and quantified. Catechins are the main compounds of flavonoids in the tea leaves, which approximately accounts for 13% of the dry weight. In summer, the amount of catechins in ‘Mooma 1’ leaves was significantly lower than the wild type (P < 0.05), whereas the amount showed no significant in spring (P = 0.08).

The O-Glycosylated flavonol biosynthesis was strengthened in the purple-leaf tea variety. The total amounts of flavonols in ‘Mooma 1’ leaves (11.26 ± 0.99 ng g⁻¹ DW in spring and 17.52 ± 1.06 ng g⁻¹ DW in summer) were notably higher than in ‘Longjing 43’ leaves (9.91 ± 0.38 ng g⁻¹ DW in spring and 12.87 ± 2.01 ng g⁻¹ DW in summer), respectively. In addition, the amounts of flavonols in summer leaves were notably higher than in spring (P < 0.05).

A representative HPLC profile of anthocyanins in ‘Mooma 1’ was presented in Fig. 2 and Table 1. Eight anthocyanins were identified in ‘Mooma 1’ leaf using LC-MS analysis. Total amount of anthocyanins in the spring leaves (May 10th, average concentration: 591.87 ± 51.1 ng g⁻¹ DW) was notably higher than in the summer (July 20th, 451.11 ± 19.02 ng g⁻¹ DW, P < 0.05), while anthocyanin concentration was very low in the wild type leaf (undetectable in two Seasons). In general, anthocyanin concentration in the purple-leaf tea was significantly higher than the wild-type cultivar, indicating that anthocyanin accumulation is responsible for the red coloration in tea.

**Comparison of transcriptional difference between purple-leaf and wild-type tea.** To deeply investigate anthocyanin and flavonol accumulation in purple-leaf tea, we carried out a comprehensive identification of transcriptional difference by RNA-Sequencing. Two normalized cDNA libraries from spring leaves of ‘Mooma1’ and wild type ‘Longjing43’ were sequenced. Following de novo assembly and redundancy reduction, we obtained a final set of 77,707 unigenes, with an average length of 1004 bp and N50 of 1675 bp (Suppl Table S2). Of these, 2,293 unigenes exhibited up-regulation, while 1,752 unigenes exhibited down-regulation in the purple-leaf transcript compared with the wild type (Suppl Figure S1). In total, 54,210 unigenes had hits in all five public databases with functional annotations (69.8% of the unigenes).

By mapping to the KEGG reference pathway, a total of 392 unigenes were assigned to the flavonoid biosynthesis pathway (including flavonoid, anthocyanin, and flavones and flavonol pathways). Among these annotated unigenes, whole sets of structural genes involved in the flavonoid biosynthesis pathway were indentified. All
flavonoid genes were multiple genes, such as 5 CsPAls, 4 CsDFRs, 3 CsCHSs, and 3 CsLARs in tea plant (Fig. 3). Unexpectedly, all the highly (FPKM > 100) and moderately (FPKM > 10) expressed genes, which are related to flavonoid biosynthesis, found no significant up-regulation in the purple variety. Only two low expressed genes, CsF3’H and CsDFR2 (FPKM < 10) showed 2- and 1.9-fold higher in purple-leaf than in green-leaf. Several CsUGTs encoding terminal enzymes by galactosylated modification of flavonoids were remarkably up-regulated in ‘Mooma1’ leaf. Especially, the expression of CsUGT72AM1 and CsUGT3 in ‘Mooma1’ leaf was increased by 4.2- and 2.5-folds respectively, compared with the value in wild-type tea. The result was further confirmed by qRT-PCR.

MYB-bHLH-WD40 (MBW) complexes, activating anthocyanin biosynthesis, were also investigated with the transcriptome of purple-leaf. The homologues of AtMYB113 (CsMYB6A, NO. CL7132), AtTT8 (CsTT8, NO. Unigene5327) and AtTTG1 (CsTTG1, NO. Unigene4594) were identified. However, CsTT8 and CsTTG1 expression levels were not significantly up-regulated in ‘Mooma1’ leaf. Only the relative expression level of CsMYB6A was significantly up-regulated in the leaves of ‘Mooma1’; and was increased by 3.3-folds compared with wild-type.

Figure 1. Phenotypic analysis of the red-leaf mutants of tea plants. (a) Wild-type tea plants and red-leaf mutants in the tea field; (b) and (c) Young branches of wild-type tea plants and red-leaf mutants; (d) and (f) Micromorphological images of the mutant leaf, ×40 fold and ×400 fold respectively; (e) and (g) Micromorphological images of the wild-type leaf staining with p-dimethylaminocinnamaldehyde (DMACA) - HCl, ×40 fold and ×400 fold respectively; Pa, palisade parenchyma; Gh, glandular hair; Xy, xylem. Note: Fig. 1: a, b and c were taken by Correspondence authors Yunsheng wang, in shitai, Chizhou, Anhui province, China, in March, 2013.
126 R2R3-MYBs from *Arabidopsis thaliana* and *CsMYB6A* were used in phylogenetic tree construction (Fig. 4). Phylogenetic analysis indicated *CsMYB6A*, attached to subgroup 6 (Sg6), was most similar to *AtMYB13* and shares 63% identity (Fig. 4). MYBs in Sg6 are reported regulators of anthocyanin accumulation28. *CsMYB6A*-GFP fusion transient expression showed that *CsMYB6A* protein was exclusively localized in the nucleus (Suppl Figure S2).

**Functional analysis of the CsMYB6A gene in Nicotiana tabacum.** A R2R3-MYB (PAP1) of *Arabidopsis thaliana* could enhance the accumulation of anthocyanins inducing the purple color of the leaf and stem29. The 35S:CsMYB6A, 35S:AtPAP1 (as positive control), and empty plasmid (as negative control) vectors were introduced into Tobacco 'G28' (*Nicotiana tabacum* 'G28'). About 10 independent transgenic tobacco plants of different genes were obtained. The leaves of *CsMYB6A* or *AtPAP1* transgenic plants exhibited a clear color change from green of the control host to purple (Fig. 5a). The concentrations of anthocyanins in the leaf of *CsMYB6A* and *AtPAP1* transgenic tobacco, as detected by HPLC method, were 240 and 340 ng g⁻¹ DW respectively, while undetected in the leaf of empty vector control (Fig. 5b). Additionally, the accumulation of flavonols in the leaf of *CsMYB6A* or *AtPAP1* transgenic tobacco was remarkably enhanced by 1.89 and 4.15-folds respectively, compared to the vector control. Environmental temperature and light are considered as important factors that affect anthocyanin biosynthesis in garden plants. Our data also confirmed that low temperature and red light accelerated the anthocyanin accumulation in the transgenic tobacco lines (Suppl Figure S3).

To investigate whether the flavonoid biosynthesis pathway was affected by over-expression of R2R3-MYBs, the flavonoid pathway genes, including *C4H*, *4CL*, *CHS*, *CHI*, *F3'H*, *F3H*, *DFR*, *FLS*, *LAR*, *ANS*, *ANR*, and *3GT* were examined by qRT-PCR in vector control and transgenic lines with β-actin (accession number: EU938079) as reference gene (Fig. 5c). The expression levels of *F3H* and *3GT* genes in the mature leaf of transgenic lines was significantly increased by 2-folds in comparison to the vector control lines, especially the level of *CHS* was enhanced by 10-folds as compared to the control lines. Therefore, the expression of these genes was stimulated by the over-expression of *CsMYB6A* or *AtPAP1* in transgenic lines. The result was also confirmed by qRT-PCR (Fig. 5c).

**Heterologous expression and enzymic analysis of the recombinant CsUGT72AM1.** RNA-Sequencing data showed that several CsUGTs, especially *CuUGT72AM1* and *CsUGT3*, were remarkably up-regulated in 'Mooma1' leaf, compared with the value in wild-type tea. Our previous research had identified a UDP-glycosyltransferase, *CsUGT78A14*, which is responsible for the biosynthesis of flavonol 3-O-glucosides30. The expression of *CsUGT78A14* was unremarkably up-regulated in the 'Mooma1’ leaf compared with wild-type control. Phylogenetic analysis, referring to *Arabidopsis* UGTs, indicated *CsUGT72AM1*, *CsUGT3* and *CsUGT78A14* are the most similar to *AtUGT72E1*, *AtUGT91A1*, and *Nt3GT*, respectively (Fig. 6).

---

**Figure 2.** HPLC chromatogram of flavonoids in wild-type tea plants and red-leaf mutants. (A), (C), and (E) Representative HPLC chromatograms of flavonoids in wild-type tea plants at 280 nm, 345 nm, and 520 nm, respectively; (B), (D), and (F) Representative HPLC chromatograms of flavonoids in red-leaf mutants at 280 nm, 345 nm, and 520 nm, respectively. Flavonoid compounds with peaks (from NO.1 to NO.24) were listed in Table 1.
### Table 1. Concentrations (mean ± SD of duplicate assays) of flavonoids in leaves of red-leaf mutant of tea plant.

| Peak number | Compound                         | Spring Green-leaf | Spring Purple-leaf | Summer Green-leaf | Summer Purple-leaf |
|-------------|----------------------------------|-------------------|-------------------|-------------------|-------------------|
|             | Catechins (mg g⁻¹DW)             |                   |                   |                   |                   |
| 1           | Gallolatechin                     | 14.30 ± 0.29b     | 9.24 ± 0.63a      | 18.16 ± 0.67c     | 18.65 ± 1.48c     |
| 2           | Epigallocatechin                  | 4.71 ± 0.31a      | 19.82 ± 1.32c     | 3.30 ± 0.17a      | 9.39 ± 8.25b      |
| 3           | Catechin                          | 1.60 ± 0.20c      | 1.29 ± 0.11b      | 0.59 ± 0.17a      | 1.12 ± 0.12b      |
| 4           | Epicatechin                       | 6.98 ± 0.44b      | 6.27 ± 0.50b      | 4.42 ± 0.65a      | 4.48 ± 0.67a      |
| 5           | Epigallocatechin gallate          | 70.31 ± 1.49b     | 56.95 ± 4.74a     | 88.72 ± 2.56d     | 77.43 ± 2.00c     |
| 6           | Epicatechin gallate               | 23.41 ± 0.93a     | 30.44 ± 2.50c     | 27.24 ± 0.94b     | 24.17 ± 0.63a     |
|             | Total                             | 124.95 ± 1.42a    | 124.00 ± 9.66a    | 142.44 ± 4.75c    | 130.56 ± 1.94b    |
|             | O-Glycosylated flavonoids (mg g⁻¹DW) |                   |                   |                   |                   |
| 7           | Myricetin-3-O-galactoside         | 0.25 ± 0.15a      | 0.44 ± 0.14ab     | 0.69 ± 0.12b      | 1.10 ± 0.14c      |
| 8           | Myricetin-3-O-glucoside           | 0.37 ± 0.09a      | 0.32 ± 0.05a      | 0.31 ± 0.03a      | 1.05 ± 0.13b      |
| 9           | Quercetin-3-O-galactosylrutinoside| 1.34 ± 0.18a      | 1.41 ± 0.10a      | 4.00 ± 0.01b      | 5.14 ± 0.88c      |
| 10          | Quercetin-3-O-glucosylrutinoside  | 0.55 ± 0.40ab     | 0.42 ± 0.02a      | 0.89 ± 0.00b      | 1.40 ± 0.25c      |
| 11          | Rutina                            | 0.41 ± 0.00c      | 0.30 ± 0.00a      | 0.57 ± 0.01d      | 0.33 ± 0.01b      |
| 12          | Kaempferol-3-O-galactosylrutinoside| 4.35 ± 0.98a     | 5.52 ± 0.30b      | 5.27 ± 0.01ab     | 5.24 ± 0.53ab     |
| 13          | Kaempferol-3-O-glucosylrutinoside | 1.04 ± 0.56a     | 1.10 ± 0.05a      | 0.73 ± 0.01a      | 1.10 ± 0.09a      |
| 14          | Kaempferol-3-O-galactoside        | 0.52 ± 0.12a      | 0.93 ± 0.05b      | 0.45 ± 0.11a      | 1.04 ± 0.10b      |
| 15          | Kaempferol-3-O-rhamnosylgalactoside| 0.69 ± 0.21a     | 0.61 ± 0.03a      | 1.14 ± 0.02b      | 0.53 ± 0.08a      |
| 16          | Kaempferol-3-O-glucoside          | 0.38 ± 0.35a      | 0.61 ± 0.03a      | 0.40 ± 0.04a      | 0.59 ± 0.03a      |
|             | Total                             | 9.91 ± 0.38a      | 11.68 ± 0.59b     | 14.43 ± 0.05c     | 17.52 ± 1.06d     |
|             | Anthocyanidins (mg g⁻¹DW)         |                   |                   |                   |                   |
| 17          | Delphinidin-3-O-galactoside       | —                  | 14.05 ± 4.06a     | —                  | 17.50 ± 1.89a     |
| 18          | Delphinidin-3-O-glucoside         | —                  | 20.70 ± 1.86a     | —                  | 21.95 ± 0.18a     |
| 19          | Cyanidin-3-O-galactoside          | —                  | 20.04 ± 1.64b     | —                  | 3.00 ± 0.13a      |
| 20          | Cyanidin-3-O-glucoside            | —                  | 1.21 ± 0.22b      | —                  | 0.51 ± 0.19a      |
| 21          | Delphinidin-3-O-coumaroylgalactoside| —                | 50.76 ± 3.20a     | —                  | 52.36 ± 6.30a     |
| 22          | Delphinidin-3-O-coumaroylglucoside| —                  | 265.24 ± 23.63a   | —                  | 222.82 ± 14.09a   |
| 23          | Cyanidin-3-O-coumaroylgalactoside | —                  | 149.94 ± 11.43b   | —                  | 122.96 ± 1.39a    |
| 24          | Cyanidin-3-O-coumaroylglucoside   | —                  | 69.93 ± 5.92b     | —                  | 10.02 ± 1.50a     |
|             | Total                             | —                  | 591.87 ± 51.16b   | —                  | 451.11 ± 19.02a   |

The enzymatic characteristics of CuUGT72A1M and CuUGT78A14 proteins that were expressed *Escherichia coli* cells were evaluated. The purified recombinant CuUGT72A1M proteins, in accordance with rCsUGT78A14, can catalyze the glycosylation of quercetin or cyanidin at 3-OH group with UDP-Glucose as the sugar donors *in vitro* (Fig. 7a,b). However, the recombinant enzyme demonstrated low activity for anthocyanins, and exhibited remarkable substrate inhibition, which was consistent with previous published reports.31

The kinetic parameters of these two recombinant glucosyltransferases for flavonols (kaempferol, quercetin, and myricetin) were determined in phosphate buffer at pH 7.5. The recombinant enzymes demonstrated highest activity for flavonols. The Km values of UGT72A1M for kaempferol, quercetin, and myricetin were 71.81, 94.99, and 113.15 µM with Vmax values of 17.92, 49.26, and 86.96 nmol min⁻¹, respectively. For UGT78A14, the Km and Vmax values of kaempferol, quercetin, and myricetin were 81.04, 72.20, and 97.35 µM and 39.84, 42.55, and 48.54 nmol min⁻¹, respectively (Fig. 7c).

### Discussion

**Flavonoid components in the leaves of purple-leaf tea.** Several health benefits were observed with tea products, which contain high concentrations of flavonoid components. This drove the scholar's initiatives to explore the specialty of tea varieties, including white tea, yellow tea, anthocyanin-rich tea. Several purple-leaf tea varieties were reported in different tea planting areas.32,33,34 A new purple-leaf tea variety ‘Mooma 1’ was reported in this paper, which was selected from the natural hybrid population of ‘Longjing 43’. The buds and young leaves of ‘Mooma 1’ displayed bright red color, where the anthocyanins were remarkably accumulated in the palisade mesophyll tissue, but were absent in the epidermis. A previous study demonstrated that the putative function of antioxidantive defense in the leaves was more possible for anthocyanins located in the mesophyll than in the epidermal vacuoles.33

Flavonoid compounds are considered to be the most important quality parameters of tea products due to their impact on color and taste properties. Catechins are responsible for astringency and bitterness.34 Flavonoids induce a velvety and mouth-coating sensation at very low concentration. Anthocyanins add to the briskness by complexing with catechins and theaflavins.35 Purple-leaf tea products show more astringency with better mouth feel and sweet after taste than green-leaf tea. Average concentrations of total anthocyanins in the spring or summer...
leaves were 591.87 ± 51.1 and 451.11 ± 19.02 ng g⁻¹ DW, respectively, which was very low in the wild type leaves. Additionally, the total amounts of O-Glycosylated flavonols in ‘Mooma 1’ leaves were significantly higher than in the control leaves.

The biosynthetic pathways of anthocyanins, flavonols, and flavan-3-ols (catechins) share common steps in the phenylpropanoid and flavonoid pathways (from PAL to F3H). Each class of flavonoid was synthesized by a multienzymatic step reaction branching from the common flavonoid pathway (Fig. 3). Several studies have shown the competitive relationship between different flavonoids due to competition for the substrate. Interestingly, our results indicated that anthocyanins and flavonols have a synergistic relationship, and both showed significant increase in the purple-leaf variety compared with the control variety. Both the compounds may have competition with catechins in purple leaves, because the total amount of catechins in purple leaves was slightly decreased compared with the control. However, further research is required to uncover the association between anthocyanin and flavonol accumulation.

An R2R3-MYB transcription factor, CsMYB6A, promotes flavonoid accumulation in purple-leaf tea. Several R2R3-MYB transcription factors, combined with other transcription factors (bHLH and WD40), are known to be involved in the regulation of flavonoid biosynthesis. AN2 from Perunia hybrid, C1 from Zea mays, and PAP1 and PAP2 from Arabidopsis thaliana all encode MYB proteins. They regulate the accumulation of different anthocyanin pigments. Pattanaik confirmed that anthocyanin related MYB interacted with other heterologous species, bHLH, to activate the expression of key flavonoid pathway genes, including CHS and DFR, which induced anthocyanin synthesis. These results indicate that regulatory anthocyanin genes are conserved between species.

Our previous study predicted R2R3-MYB genes in flavonoid biosynthesis of C. sinensis, while Sg6 CsMYB genes were lacked in the wild-type green-leaf tea. In the current study of ‘Mooma 1’ purple-leaf cultivars, we
found that the expression of the CsMYB6A was most similar to AtMYB113, and was consistent in regulating anthocyanin synthesis. The different expression levels of CsMYB6A in different cultivars might be due to the transcriptional modification or post-transcriptional modification. Furthermore, we also investigated the over-expression of CsMYB6A and AtPAP1 genes in transgenic tobacco, G28. Results showed that there was significant up-regulation of structural flavonoid genes, especially CHS and anthocyanin 3-O-glucosyltransferase (A3T) in both CsMYB6A and AtPAP1 transgenic tobacco lines compared with empty-vector control. Interestingly, the expression levels of CsUGT78A14, the homologous gene of 3GT, demonstrated no significant up-regulation in the leaves of ‘Mooma1’, while several other UGTs, especially CsUGT72AM1, were obviously up-regulated compared with green-leaf tea. These results indicated that the divergent target genes were responsible for the species-specific differences in regulatory networks.

CsUGT72AM1, an UGT gene, catalyzes the glycosylation of flavonoids. UDP-glycosyltransferases are the final enzymes produced in anthocyanin and flavonol biosynthesis. 78 UGT family genes, such as UGT78G1, UGT78D2 were activated with anthocyanidins, flavonols, flavones, coumestans, pterocarpans, and isoflavones, and were involved in the O-glucosylation of anthocyanins. In our previous study, CsUGT78A14 was found to be responsible for the biosynthesis of flavonol 3-O-glucosides. A novel UGT gene, CsUGT72AM1, showed higher expression level in ‘Mooma1’, and was screened and cloned from purple-tea leaves. Phylogenetic study indicated that CsUGT72AM1, which is similar to AtUGT72E1, was clustered into the 72 UGT family of glucosyltransferases. The 72 UGT subgroup, studied in several previous studies, was able to catalyze the formation of monolignol 4-O-glucose that is involved in the biosynthesis of lignin.
Several UGT family genes can catalyze the O-glucosylation of flavonoids. For example, UGT72L1 from Medicago truncatula, which was regulated by Arabidopsis R2R3-MYB (TT2) was specifically active towards the epicatechin, formatting epicatechin 3′-O-glucoside42.

Figure 5. Phenotypic analysis of transgenic tobacco overexpressing CsMYB6A and AtPAP1. (a) Leaves of transgenic plants (CsMYB6A and AtPAP1) in comparison with empty-vector transgenic plants; (b) Total concentrations of anthocyanins and flavonols of transgenic plants. The relative flavonol concentration was calculated as the ratio between the total peak area at 350 nm; (c) The relative gene expression involved in flavonoid biosynthesis in transgenic plants compared with the empty vector. Relative FPKM and expression data were obtained from RNA-sequencing and qRT-PCR analysis, respectively.
In this paper, we compared the enzymatic characteristics of \( Cs\text{UGT72AM1} \) and \( Cs\text{UGT78A14} \) through \textit{in vitro} assays. Our enzymatic assays confirmed that both \( Cs\text{UGT72AM1} \) and \( Cs\text{UGT78A14} \) demonstrated catalytic activity as a flavonol or anthocyanin 3-O-glucosyltransferase. The recombinant enzymes \( Cs\text{UGT72AM1} \) and \( Cs\text{UGT78A14} \) were highly catalyzed by the addition of glycosyl group from UTP-glucose to flavonols. \( Cs\text{UGT72AM1} \) displayed broad substrate specificity, by \textit{in vitro} experiments, recognizing the flavonoid substrates, including naringenin (N), Eriodictyol (E), kaempferol (K), quercetin (Q), and myricetin (M), as acceptor molecules (Data not shown in this paper). These results indicated that both \( Cs\text{UGT72AM1} \) and \( Cs\text{UGT78A14} \) genes are likely to be involved in the biosynthesis of flavonoid 3-O-glycoside compounds in tea plants.

Taken together, our research acquired the key genes, including \( Cs\text{MYB6A} \) and \( Cs\text{UGT72AM1} \), which regulated the accumulation of anthocyanin and flavonol in purple-leaf tea. \( Cs\text{MYB6A} \) transgenic results indicated that the regulatory mechanism of flavonoid pathway was conserved between species\(^5\). In future, we would focus on exploiting MYB-target genes using interaction methods. Therefore, \( Cs\text{MYB6A} \) and related \( Cs\text{UGTs} \) are the first characterized genes that are involved in the anthocyanin biosynthesis in tea leaf organs. Their identification provides advances in understanding the flavonoid biosynthesis in non-alcoholic beverage crops.

Materials and Methods

Plant material and standard chemicals. Leaves of tea variety ‘Mooma 1’ and wild type ‘Longjing 43’ (Camellia Sinensis CV ‘Longjing’) were sampled from healthy pants that are grown in the tea garden of Shitai County, Anhui, China (latitude 30.19 N, longitude 4 E, altitude 20 m above mean sea level). A total of 10 buds

---

**Figure 6.** Phylogenetic tree of UGT proteins. The protein sequences contained \( Cs\text{UGT72AM1} \) (accession number: KY399734), \( Cs\text{UGT78A14} \) (accession number: KP682360), \( Cs\text{UGT3} \) from \textit{Camellia sinensis} and \( At\text{UGTs} \) from \textit{Arabidopsis thaliana} (http://www.arabidopsis.org/browse/genefamily/index.jsp). The phylogenetic tree was constructed using MEGA 5 with 1000 bootstrap replicates. Number indicates the percentage of consensus support.
and leaves were randomly collected from different branches, frozen immediately in liquid nitrogen, and stored at −80 °C for RNA-seq, qRT-PCR and HPLC analysis.

Catechin (C), gallocatechin (GC), epicatechin (EC), epigallocatechin (EGC), epicatechin 3-O-gallate (ECG), epigallocatechin 3-O-gallate (EGCG) were obtained from Shanghai RongHe Pharmaceutical Co. Myricetin 3-O-glucoside, quercetin 3-O-glucoside, kaempferol, petunidin, cyanidin, cyanidin 3-O-glucoside, and delphinidin were purchased from Sigma Chemicals Co.

Tissue slicing and staining. Samples for observation were prepared by standard free-hand sectioning. To slice the tea tissues, fresh carrot was used as a supporter. The section was observed under a microscope (Olympus, Tokyo, Japan). The sections were stained using 0.01% (w/v) 4-dimethylaminocinnamaldehyde (DMACA) in absolute ethanol containing 0.8% w/v hydrochloric acid to observe the presence of catechins in the tissue of tea leaf.

Analysis of catechins, flavonols, and anthocyanins in leaves. Leaves were ground to a fine powder in liquid nitrogen. The powder (1 g) was extracted with 5 ml methanol at room temperature for 10 min, followed by centrifugation at 4000 × g for 15 min. The residue was re-extracted thrice by this method. The supernatants were filtered through a 0.22 μm membrane. Catechins, flavonols, and anthocyanins were analyzed according to the liquid chromatography–mass spectrometry (LC-MS) and HPLC methods. Catechins, flavonols, and anthocyanins were quantified at 280 nm, 345 nm, and 530 nm respectively. Since only 11 standards were available, myricetin 3-O-glucoside was used as the molar equivalent to quantify its derivatives, quercetin 3-O-glucoside for all the quercetin 3-O-glycosides, kaempferol for kaempferol 3-O-glycosides, petunidin for petunidin 3-O-glycosides, cyanidin for cyanidin 3-O-glycosides, and delphinidin for delphinidin 3-O-glycosides. All samples were run in triplicate for both quantitation and multivariate statistical analysis.

cDNA library construction and transcriptome sequencing. Total RNA of tea and tobacco (Nicotiana tabacum) leaves was isolated with RNAiso Plus and RNAiso-mate for Plant Tissue kits and treated with DNase I according to manufacturer's instructions (Takara, China). RNA quality was examined using 1% agarose gel and the concentration was determined using a Nanodrop spectrophotometer (Thermo, Waltham, MA, USA). Illumina sequencing was performed at Beijing Genomics Institute (BGI, Wuhan, China) on the HiSeq™ 2000 platform (Illumina, San Diego, CA) and the bioproject ID was PRJNA283232 in NCBI (https://www.ncbi.nlm.nih.gov/bioproject/?term). The de
novel assembly, functional annotation, and metabolic pathway analysis were carried out by BGI Institute according to the manufacturer's instructions. Genes involved in flavonoid pathway were further confirmed using BLASTn and BLASTp tools in Tea tree Genome Database (http://www.plantkingdomgdb.com/tea_tree/). The NCBI accession numbers of the flavonoid-related Genes were listed in Table S1.

**qRT-PCR analysis.** Quantitative real-time PCR (qRT-PCR) was performed using the SYBR® Premix Ex Taq™ II (Perfect Real Time) kit (Takara, Japan) on a Peltier Thermal Cycler PTC200 (Bio-Rad, USA) with gene-specific primer pairs (Table S1). The related expression level was normalized against the expression level of the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in tea plant48 or actin in tobacco49. The melting curve was performed to determine the PCR product size and to detect possible primer dimers. Triplicates of all samples were run. The cycle number at which the reaction crossed an arbitrarily placed threshold (CT) was determined for each gene, and the relative expression of each gene was determined using the equation

\[ \Delta \Delta CT = (C_{\text{Target}} - C_{\text{GAPDH}})_{\text{Sample}} - (C_{\text{Target}} - C_{\text{GAPDH}})_{\text{Control}} \]

**Transformation of tobacco plants with CsMYB6A.** The Gateway Cloning System was used to construct the transformation vectors of CsMYB6A. The PCR primer pairs for linking the attB adaptors are listed in the Additional file 2: Suppl Table S1. CsMYB6A PCR product was cloned into the entry vector pDONR207 by Gateway BP Clonase Enzyme mix according to the manufacturer’s instructions (Invitrogen, USA). The pDONR207-CsMYB6A entry vector was then transferred into the Gateway plant transformation destination vector pCB2004 using Gateway LR Clonase (Invitrogen, USA). Recombinant colonies pCB2004-CsMYB6A and control pCB2004 vectors were selected on kanamycin plates and validated by bacterial colony PCR, followed by transformation into EHA105 by electroporation at 2500 V for 5.5 ms. A single colony containing each target construct was confirmed by PCR and then used for genetic transformation of tobacco. EHA105-pCB2004-CsMYB6A and EHA105-pCB2004 were prepared for transformation. The leaf disc approach was used for tobacco transformation with 25 mg/L phosphinothricin selection.

**Expression and purification of recombinant CsUGT72AM1.** The cDNA of CsUGT72AM1 were sub-cloned into the expression vector pMAL-c2X (New England Biolabs, MA, USA). The cloned gene sequences were also confirmed by colony PCR. The pMAL-CsUGT72AM1 expression vector sand empty vectors were transformed into E. coli Novablate (DE3) competent cells (Novagen, Schwabach, Germany). Recombinant proteins were purified according to the manufacturer's instructions (New England Biolabs, MA, USA). Recombinant enzyme assays were carried out as described by Cui et al.50. The Km and Vmax of CsUGT72AM1 was determined using 5 mM UDP-glucose (UDP-Glc) as the sugar donor and 1.5–200 μM of flavonols as acceptors (kaempferol and quercetin) in phosphate buffer (pH 7.5). Reaction samples lacking recombinant proteins were used as blank controls. Reactions were stopped by mixing the reaction solutions with 100% methanol. All the kinetic assays were incubated at 30 °C for 10 min and repeated in triplicate.

**Bioinformatics and statistical analyses.** Multiple sequence alignment was performed using ClustalX (http://www.clustal.org). Phylogenetic tree was constructed using protein sequences from several plant MYBs and UGTs by Neighbor-Joining distance analysis using MEGA5.0 (http://www.megasoftware.net/). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the p-distance method.

Data were presented as mean ± SD. Statistically significant differences between the groups were determined with Student’s t-test using SPSS software (SPSS, Chicago, IL, USA). P < 0.05 was considered to be statistically significant.

**References**

1. Jasiński, M., Kachlicki, P., Rodziewicz, P., Figlerowicz, M. & Stobiecki, M. Changes in the profile of flavonoid accumulation in Medicago truncatula leaves during infection with fungal pathogen Phoma medicaginis. *Plant Physiology & Biochemistry* 47, 847–853 (2009).
2. Li, J., Ou-Lee, T.-M., Raba, R., Amundson, R. G. & Last, R. L. Arabidopsis flavonoid mutants are hypersensitive to UV-B irradiation. *The Plant Cell* 5, 171–179 (1993).
3. Azuma, A., Yakushiji, H., Koshita, Y. & Kobayashi, S. Flavonoid biosynthesis-related genes in grape skin are differentially regulated by temperature and light conditions. *Planta* 236, 1067–1080 (2012).
4. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* 126, 485–493 (2001).
5. van Tunen, A. J., Mur, L. A., Recourt, K., Gerats, A. & Mol, J. Regulation and manipulation of flavonoid gene expression in anthers of petunia: the molecular basis of the Po mutation. *The Plant Cell* 3, 39–48 (1991).
6. Jennings, A. et al. Higher anthocyanin intake is associated with lower arterial stiffness and central blood pressure in women. *American Journal of Clinical Nutrition* 96, 781 (2012).
7. Chang, J. et al. Mulberry anthocyanins inhibit oleic acid induced lipid accumulation by reduction of lipogenesis and promotion of hepatic lipid clearance. *Journal of Agricultural and Food Chemistry* 61, 6069 (2013).
8. Rahman, M. M., Ichiyangagi, T., Komiyama, T., Sato, S. & Koshioka, T. Effects of anthocyanins on psychological stress-induced oxidative stress and neurotransmitter status. *Journal of Agricultural and Food Chemistry* 56, 7545–7550 (2008).
9. Hidalgo, M. et al. Potential anti-inflammatory, anti-adhesive, anti-estrogenic, and angiogenesis-converting enzyme inhibitory activities of anthocyanins and their gut metabolites. *Genes & nutrition* 7, 295–306 (2012).
10. Kuo, C.-Y. et al. Hibiscus sabdariffa L. extracts reduce serum uric acid levels in oxonate-induced rats. *Journal of Functional Foods* 4, 375–381 (2012).
11. Wang, Y. & Ho, C.-T. Polyphenolic chemistry of tea and coffee: a century of progress. *Journal of Agricultural and Food Chemistry* 57, 8109–8114 (2009).
12. Joshi, R., Rana, A. & Gulati, A. Studies on quality of orthodox teas made from anthocyanin-rich tea clones growing in Kangra valley, India. *Food Chemistry* 176, 357–366 (2015).
31. Kovinich, N., Saleem, A., Arnason, J. T. & Miki, B. Functional characterization of a UDP-glucose: flavonoid 3-O-glucosyltransferase.

32. Kerio, L. C., Wachira, F. N., Wanyoko, J. K. & Rotich, M. K. Total polyphenols, catechin profiles and antioxidant activity of tea plants from the seed coat of black soybean (Glycine max (L.) Merr.

34. Kallithraka, S., Bakker, J. & Clifford, M. Evaluation of Bitterness and Astringency of (+)-Catechin and (-)-Epicatechin in Red Wine and in Model Solution.

39. Li, Y.-Y.

51. Lei, Z. Y.

26. Zhao, L.

22. Nesi, N., Jond, C., Debeaujon, I., Caboche, M. & Lepiniec, L. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed.

24. Terrier, N. et al. Ectopic expression of VmMYBPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway.

27. Zhao, L. et al. The R2R3-MYB, HHLH, WD40, and related transcription factors in flavonoid biosynthesis. Functional & integrative genomics 13, 75–98 (2013).

30. Cui, L.

45. Liu, Y., Gao, L., Xia, T. & Zhao, L. Investigation of the site-specific accumulation of catechins in the tea plant (Camellia sinensis (L.).

46. Abeynayake, S. W., Panter, S., Mouradov, A. & Spangenberg, G. A high-resolution method for the localization of proanthocyanidins in plant tissues.

47. Wang, Y.

50. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2− (ΔΔC(T)) Method.

52. Consonni, G., Guna, F., Gavazzi, G. & Tonelli, C. Molecular homology among members of the R gene family in maize.

58. Tian, Y., Jia, Y. et al. Anthocyanins from new red leaf tea ‘Sunrouge’. Plant physiology 165, 2293–2306 (2014).

60. Lei, Z. Y. et al. High-throughput Binary Vectors for Plant Gene Function Analysis. Journal of Integrative Plant Biology 49, 556–567 (2007).

63. Tsuda, Y., Kato, T., Inada, M. & Tanaka, S. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. New phytologist 153, 387–400 (2001).
Acknowledgements
We thank the Chengbi Xiang Lab. (University of Science and Technology of China, Hefei city, China) for excellent assistance transgenic expression in tobacco leaves. We also thank Huarong Tan for assistance with the LC/MS analysis. This work was supported by the Natural Science Foundation of China (31770729, 31570694, 31470689, 31300577, 31270730, and 31200229), the Postgraduate Foundation of Anhui Agricultural University, Anhui Province China (2017yjs-36), the Natural Science Foundation of Anhui Province (1708085QC51), the Collegiate Natural Science Foundation of Anhui Province (KJ2016A223), the Natural Science Foundation of Anhui Province, China (1408085QC51), the Special Foundation for Independent Innovation of Anhui Province, China (13Z03012).

Author Contributions
Conception and designed the study: Yunsheng Wang, Tao Xia, Xiujuan He. Drafted the manuscript: Yunsheng wang, Xiujuan He, Xuecheng Zhao. Performed the experiments: Xuecheng Zhao, Xinlong Dai, Xingxing Shi, Yajun Liu, Xiujuan He. Analyzed the data: Liping Gao, Xiujuan He, Xuecheng Zhao, Yajun Liu. Contributed reagents/materials/analysis tools: Liping Gao, Xiaolan Jiang, Xiujuan He, Xingxing Shi, Xinlong Dai, Yajun Liu. All authors read and approved the final manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-18133-z.

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017