The functional analysis of the dihydroflavonol 4-reductase family of *Camellia sinensis*: exploiting the key amino acids to reconstruct the reduction activity

Haixiang Ruan\textsuperscript{1,2#}, Xingxing Shi\textsuperscript{1,3#}, Liping Gao\textsuperscript{2#}, Arif Rashid\textsuperscript{2}, Yan Li\textsuperscript{2}, Ting Lei\textsuperscript{2}, Xinlong Dai\textsuperscript{3}, Tao Xia\textsuperscript{1*}, and Yunsheng Wang\textsuperscript{1,2*}

\textsuperscript{1}State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, Anhui 230036, China; \textsuperscript{2}School of Life Science, Anhui Agricultural University, Hefei, Anhui 230036, China; \textsuperscript{3}College of Tea Science, Guizhou University, Guiyang Guizhou 550025, China;
\# These authors contributed equally to this work.
* Correspondence: wangyunsheng@ahau.edu.cn; xiaota62@126.com

**Running title:** Key amino acid controlling DFR reductase activity

**Keywords:** *Camellia sinensis*, dihydroflavonol 4-reductase, dihydroflavonols, enzymatic characteristics, metabolic engineering, site-directed mutagenesis

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. This is an Open Access article distributed under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Highlight: We reported the first dihydroflavonol 4-reductase (CsDFRs) class II/III genes in plants. This not only contributes to the DFR gene evolution, but provides candidate DFR gene for flavonoid metabolic engineering.

Abstract:

Anthocyanin and proanthocyanidins (PAs) are important components of flavonoids, secondary metabolites in plants with a wide range of industrial and pharmaceutical applications. DFR (dihydroflavonol 4-reductase) is a pivotal enzyme which plays an important role in the flavonoid pathway. Here, four CsDFRs genes were isolated from Camellia sinensis and their overexpression were analyzed in vitro and in vivo. Based on transcription and metabolic analyses, the CsDFRs expression is closely consistent with catechins and PAs accumulation. Moreover, the enzyme activity analyses revealed that the two recombinant proteins CsDFRa and CsDFRc exhibited DFR activity, which converts dihydroflavonols into leucoyanidin in vitro, but not CsDFRb1 and CsDFRb3. CsDFRa and CsDFRc overexpression in AtDFR mutants (tt3) revealed that CsDFRs are involved in the biosynthesis of anthocyanins and PAs, as CsDFRa and CsDFRc not only restored the petiole purple phenotype, but also the seed coat color. Site-directed mutagenesis revealed that the two amino acid residues S117 and T123 of CsDFRa play a prominent role in controlling DFR reductase activity. Enzymatic assays indicated that CsDFRa and CsDFRc exhibited a higher affinity for DHQ and DHK, respectively, whereas CsDFRb1N120S and
CsDFRb1^{C126T} exhibited a higher affinity for DHM. Our findings present comprehensive characterization of the DFRs from C. sinensis and shed light on their critical role in metabolic engineering.

**Introduction**

Flavonoids, which include flavones, flavanols, flavonols, flavan-3-ols (catechins), anthocyanidins and proanthocyanidins (PAs), are a class of naturally occurring secondary metabolites distributed widely in different tissues of plants, such as leaves, flowers, stems, and fruits. Flavonoids are also the most potent and health-promoting compounds. In epidemiological, clinical, and animal studies, they fight against different diseases, such as cardiovascular disease, cancer, and other disorders. The metabolic and biosynthesis pathways of plant flavonoids have piqued interest in the field of plant secondary metabolites due to their functionalities, compelling evidence of health benefits, industrial applications, and incorporation into various food products.

The flavonoid biosynthesis pathway in numerous plant species has been extensively investigated at the genetic, molecular and biochemical levels, with the identification of key functional genes and transposable elements having been identified. Naringenin is an important precursor that can be hydroxylated by flavonoid-3-hydroxylase (F3H) at the 3-position of the C ring or hydroxylated by flavonoid-3’-hydroxylase (E3H) or flavonoid-3’5’-hydroxylase (F3’5’H) to produce three types of dihydroflavonols, including DHK (dihydrokaempferol), DHQ (dihydroquercetin) and DHM (dihydromyricetin). DFR (dihydroflavonol 4-reductase) is an enzyme that converts dihydroflavonols to leucoanthocyanidins, which are precursors to anthocyanidins, flavan-3-ols, and PAs (Figure 1A).
Figure 1. The flavonoid biosynthesis pathway and the end-product accumulation of flavonoids in *C. sinensis*. (A) The flavonoid biosynthesis. DFR, dihydroflavonol 4-reductase; CHI, chalcone isomerase; ANS, anthocyanidin synthase; CHS, chalcone synthase; F3'H, flavonoid 3'-hydroxylase; F3H, flavanone 3-hydroxylase; F3'5'H, flavonoid 3',5'-hydroxylase; FLS, flavonol synthase; LAR, leucoanthocyanidin reductase; ANR, anthocyanidin reductase; (B) The proportion of various flavonoid compounds. The proportion of flavonoid compounds were taken from Jiang.

DFR is a flavonoid biosynthesis key rate-limiting enzyme that belongs to the NADPH-dependent epimerase/dehydratase family. The genes that code for the DFR enzyme have been characterized and isolated from a variety of plant species, and their functions have been extensively studied. Several reports have shown that *DFR* is primarily involved in the color of the petals and seed coat. The *Arabidopsis* DFR-deficient mutant (*tt3*), for example, has a complete lack of PAs accumulation and the seed coat becomes transparent, and the DFR-deficient gerbera has a normal phenotypic of white-colored petals. Other *DFR*-like genes in plants, such as *Oryza sativa*'s *OsDFR2* and *DRL1* from *Arabidopsis*, have been established to have various biochemical and physiological activities in plants, and are essential for pollen formation and male fertility.

Several studies have shown that dihydroflavonol 4-reductases from different species have distinct
substrate specificity, such as the DFRs from *Cymbidium* and *Petunia*, which did not effectively reduce DHK, the precursor to orange pelargonidin-type anthocyanins. Previously, a 26 amino acid region in the middle of the sequence was identified as determining the substrate specificity of DFR proteins.

*C. sinensis* is a commercially significant plant that produces one of the most popular non-alcoholic beverages on the planet. Besides its distinct flavor, tea is regarded as a health drink due to its nutritional and therapeutic properties. Tea has become one of the most prominent symbols of Chinese culture. Tea health promoting properties have been mainly attributed in recent years to its polyphenol content, particularly flavan-3-ols (catechins) and flavonols, which comprise 30% of fresh leaf dry weight. Catechins with gallic acid esters, such as epigallocatechin-3-gallate (EGCG) as well as epigallocatechin (ECG), are the most abundant catechins in tea and have gained the most attention. The total amount of PAs in the root, particularly proanthocyanidins dimers and trimers, however, significantly higher than in the leaves and stems (Figure 1B and Table S1). Overall, these findings reveal that the tea plant accumulates a high amount of catechins, but the molecular mechanism underlying catechin metabolism and regulation in tea leaves remains elusive.

Previous studies indicate that *CsDFRa* is a key gene in regulating catechin content in tea plants. Previously, five *CsDFR* genes have been identified using genome sequence and transcriptome databases. In this study, four *CsDFR* genes (*CsDFRa, CsDFRb1, CsDFRb3, CsDFRc*) were identified and cloned from the bud/root of *C. sinensis* cv. Shuchazao, and the expression patterns in different tissues were analyzed using qRT-PCR. A phylogenetic analysis was performed which revealed that CsDFRa was clustered into class I, and the other three proteins of *C. sinensis* were clustered independently outside the other NADPH-dependent reductase branches (class II and class III). The prokaryotic expression vector was constructed and the enzyme activity was determined in vitro. Moreover, to elucidate the function of CsDFRs in vivo, CsDFRs were overexpressed in the *Arabidopsis tt3* mutant to observe phenotypic and metabolic changes. To determine the region of the DFRa and DFRb1 enzyme responsible for the substrate specificity, we introduced chimeric DFRs and determined the activity region. Furthermore, we confirm that the
activity of DFR can be modulated by the alteration of the amino acids in the region. This paper reports the functional identification of DFR class II and class III genes in plants comprehensively and systematically, and it is proposed that these findings will not only aid in the study of the DFR gene family evolution of in shrub/vine plants, but will also provide new candidate DFR genes for flavonoid metabolic engineering.

Results

Cloning and sequence analysis of CsDFRs

Based on the genome sequence and transcriptome information of C. sinensis, five putative CsDFRs genes were screened and identified, including CsDFRa (GeneBank ID: KY615690), CsDFRb1 (KY615691), CsDFRb2 (KY615692), CsDFRb3 (KY615693), and CsDFRc (KY615694). To clone the CsDFR genes, cDNA libraries were constructed from the buds and roots of cv. Shuchazao and screened using the gene-specific primers of CsDFRs from C. sinensis as listed in Table S2. The other four genes, except CsDFRb2, have been successfully cloned and sequenced (Figure S1A). The sequencing results showed that the ORF of CsDFRa, CsDFRb1, CsDFRb3 and CsDFRc are 1044bp, 1035bp, 1023bp and 1074 bp, respectively, encoding 347, 344, 340 and 357 amino acids (Figure S1A). The characteristics of four structural genes were shown in Figure S1B. The four DFR genes have similar genetic structures with 6 exons and 5 introns, similar to DFR genes from other plants such as Arabidopsis and buckwheat. To understand further the function of the DFR genes, we predicted the major transcription factor binding motif in the promoter region (Figure S1C), and MYB binding sites have been detected in the promoter region of DFRa, DFRb1, and DFRc. However, no MYB binding site was detected in the promoter region of DFRb3. In the promoter region of DFRb1, the MYB motifs MYBPLANT and MYBPZM were not found.

Sequence alignment with several identified DFRs revealed that the N-terminus regions of the C. sinensis DFR proteins comprised putative NADPH and substrate binding sites (Figure 2A). Compared with other DFR proteins, the substrate binding regions of the CsDFRb3 and CsDFRc sequences have 5 or 3 redundant residues, respectively. DFR proteins can be classified into three
types based on the conservation of the 134th residue: asparagine (Asn), aspartic acid (Asp) and neither-Asn nor Asp. CsDFRa, like most DFR proteins from other species, is an Asn-type protein whereas the other three proteins from *C. sinensis* are neither Asn nor Asp-type (Figure 2).

A phylogenetic tree was constructed using the neighbor-joining method to further investigate the homology of these four proteins with other known DFRs and NADPH-dependent reductases, such as anthocyanidin reductase (ANR), leucoanthocyanidin reductase (LAR), and anther-specific DFR-like enzymes. The result showed that CsDFRa was clustered into class I, with proteins from dicotyledon, monocotyledon, and gymnosperm species. In class I, several proteins have been found in previous literature (Figure 2B). CsDFRa was most similar to *Actinidia chinensis* (identified score: 98.86%), *Rhododendron simsii*, and *Vaccinium ashei* (89.28%). Furthermore, the other three proteins of *C. sinensis* were clustered independently outside the other NADPH-dependent reductase branches (class II and III), and CsDFRb1 and b3 were slightly closer to the core DFR branch. To our knowledge, the functions of genes in classes II and III have not been confirmed *in vivo* or *in vitro*, and most of the genes are from shrubs or vines, especially from *Actinidia chinensis*, *Vitis vinifera*, and *Manihot esculenta*. Therefore, CsDFRb1, b3, and CsDFRc might have special significance for studying the evolution of DFR gene family. The function of these genes needs to be further elucidated.
Figure 2. Sequence analysis of CsDFRs. (A) Multiple sequence alignment of DFRs from various plant species. The DFR featured motifs, including the NADPH binding region and substrate binding site are boxed in red and blue, respectively. (B) Phylogenetic tree of CsDFRs and other reductase different plant species. The accession numbers were obtained from publically accessible database NCBI, and are presented in the figure.

Expression pattern of CsDFRs in tea plants

Quantitative RT-PCR was used to detect the expression patterns of CsDFR genes in tea plants. The gene-specific primers were used for the amplification of the expression tendency of four genes in distinct tissues (buds and young leaf, mature leaf, stem and root) (Figure 3A). For normalization,
the gene *GAPDH* was used as a reference control. Each gene was differentially expressed in distinct tissues (Figure 3B), such as transcripts level of *CsDFRa* were detected in all tissues examined, with the highest expression level in the bud, followed by the young leaf and stem, and a low expression level was detected in the mature leaf and root. The expression pattern of *CsDFRa* is consistent with the concentration of catechins in tea plants (Figure 1B). However, the expressions of other genes, such as *CsDFRb1*, *CsDFRb3*, and *CsDFRc*, were highest in the root and lowest in the mature leaf and stem, which is consistent with the concentration of PA and indicating the involvement of these three genes in the PA biosynthesis in tea plant. Moreover, based on RNA sequencing and qRT-PCR results, the expression level of *CsDFRa* (the mean Ct: 19.67, the RPKM: 696.54 in the bud) was significantly superior than the other three genes (Dataset S1).

![Figure 3. Temporal-spatial expression pattern of CsDFRs genes in C. sinensis using Quantitative RT-PCR analysis](https://academic.oup.com/hr/advance-article/doi/10.1093/hr/uhac098/6572277)

Figure 3. Temporal-spatial expression pattern of CsDFR genes in *C. sinensis* using Quantitative RT-PCR analysis (A) the figure shows different tissues collected from numerous plants for gene expression. (B) Transcripts levels CsDFRs. The data are represented as a mean and ± standard deviation of three biological replicates (*n* = 3), and *GAPDH* were used as internal control for the normalization of the Ct values. The alphabets (a, b, c, d, e and f) represents the significance level at *P*<0.05.

**Recombinant CsDFRs proteins catalytic activity in vitro**

To analyze further the enzymatic properties of CsDFRa, CsDFRb1, CsDFRb3, and CsDFRc, a series of SUMO vectors carrying different *CsDFR* genes were constructed and successfully expressed in *E. coli*. The SDS-PAGE electrophoretogram showed that clear target bands could be
obtained after IPTG inducement of recombinant strains (Figure 4A).

We detect the product anthocyanin, splitted from the leucoanthocyanidin under high temperature and acidic conditions, because the leucoanthocyanidin produced by DFR enzyme is highly unstable and difficult to detect directly by using HPLC. The red products were observed in the culture of the recombinant strain carrying the CsDFRa or CsDFRc gene, using different substrates, including DHK, DHQ, or DHM. As illustrated in Figure 4B, when dihydroflavonols were used as a substrate, no red products were observed, indicating that CsDFRb1 and b3 had little or no reductive activity.

Figure 4. Analyses of catalytic reaction product with CsDFRs fusion proteins. (A) Schematic map of the recombinant CsDFR construct and the profiles of expressed CsDFRs fusion protein. M: Protein molecular weight markers; 1: Before induction; 2: The total proteins of induction by SUMO-CsDFRs; 3: Purified protein. (B) Analyze of the enzyme reaction with SUMO-CsDFRs in vitro, and DHK, DHQ and DHM as the substrate.
Functional confirmation of the CsDFRs in Arabidopsis

In A. thaliana, tt3 was the DFR genes mutant, which not only showed a deficiency in anthocyanin synthesis but also exhibited an absence of the seed coat pigment. Therefore, the tt3 mutant is a suitable model plant for determining whether CsDFR is associate with the biosynthesis of anthocyanin and PAs.

The ORFs of three CsDFRs (including CsDFRa, CsDFRb1, and CsDFRc) were transformed into tt3 mutant driven by 35S promoter, to elucidate whether CsDFRs have a functional tt3 ortholog that can restore the above-mentioned deficient phenotypes of Arabidopsis tt3 mutant. After overexpression of CsDFRs in Arabidopsis tt3 mutant, perceptible phenotypic vary were observed in the CsDFRa and CsDFRc transgenic seedlings and/or seed coats relative to the tt3 mutant (Figure 5A). The CsDFRa and CsDFRc not only restored the petiole purple phenotype but also recovered the seed coats color of the tt3 mutant, as compared to the tt3 mutant and CsDFRb1 transgenic lines (Figure 5A).

DFR converts dihydroflavonols into leucoanthocyanins in the flavonoid biosynthesis pathway. Following that, leucoanthocyanins were further converted into anthocyanins by ANS. These aforementioned compounds mentioned will play a role in the synthesis of PAs in plants. Polyphenols were extracted from wild type (Ler), tt3 mutant, and CsDFRs transgenic lines to detect the contents of anthocyanins and soluble PAs in the seedlings and seeds. CsDFRa and CsDFRc transgenic Arabidopsis had significantly higher anthocyanins content in the seedlings than CsDFRb1 and tt3 mutant lines (Figure 5B). Furthermore, DMACA staining revealed that bluish compounds were formed in the seeds extract of the CsDFRa and CsDFRc transgenic lines, indicating that the PAs were increased in the seeds compared to the control tt3 mutant, along with wild type as a positive control (Figure 5B). These findings suggested that CsDFRa and CsDFRc were involved in the biosynthesis of anthocyanin and PAs, and they are consistent with the catalytic function of CsDFRa and CsDFRc, which can catalyze the dihydroflavonols into leucoanthocyanidins, but CsDFRb1 cannot in vivo.
Figure 5. Overexpression CsDFRα in Arabidopsis mutant tt3. (A) The phenotype of the Ler, tt3, CsDFRα/tt3, CsDFRβ1/tt3 and CsDFRγ/tt3 in seedling or seeds. (B) The absorbance values of the anthocyanin extracts and of bluish compounds from DMACA reactive compounds at 530 nm and 640 nm, respectively. All the values are the means of three biological replicates, and the different letters (A, B, C and a, b, c) indicated the significant level at $P < 0.05$.

Identification of a region that determines the activity defereence between CsDFRα and CsDFRβ1

Although the amino acid sequences of the CsDFRβ1 and CsDFRα are quite similar, the CsDFRα can reduce dihydroflavonol, while the CsDFRβ1 cannot. We used cDNA sequences from these two genes to create five chimeric DFR genes in order to find the area that controls the activity difference between CsDFRα and CsDFRβ1. (Figure 6A and Dataset S2). Prior to catalytic analysis, the purified proteins were prepared and purified. The reducing activity of DHQ was used as the substrate to test the reducing activity of these chimeric DFR proteins (Figure 6B).
Figure 6. Truncation and reorganization that determines the putative activity switch region between CsDFRa and CsDFRb1. (A) The model of amino acid position by truncation. FS1, FS2, FS3, FS4 and FS5 stands for five different reorganization strategies, amino acid residues with red represent fusion between CsDFRa and CsDFRb1, substrate-binding region and putative activity switch region are boxed in black and blue, respectively. (B) Analyze of the enzyme reaction with CsDFRa, FS1, FS2, FS3, FS4, FS5 and CsDFRb1 in vitro and DHQ as the substrate.

The chimeric gene FS1 was constructed with 5’-CsDFRb1 and 3’-CsDFRa at the start of the CsDFRa substrate-binding region, while FS2 was created with 58 amino acids before the substrate-binding region. The results of activity tests revealed that cyanidin formation could be observed in the reaction system with FS1 protein in vitro, but not in the reaction system with FS2 protein. We speculate that the DFR protein has an activity switch region prior to the substrate-binding region. Based on this speculation, three chimeric genes were further constructed, and the sequence of CsDFRb1 was gradually being replaced by CsDFRa before the substrate-binding region. The activity test results showed that the FS3 and FS4 recombinant proteins exhibited reductase activity, whereas the FS5 recombinant enzyme has lost its reductase activity. Therefore, it was predicted that the amino acid residues between FS4 and FS5 recombination points might have significant impact on DFR enzyme activity.

Site-directed mutagenesis of the activity switch region

The amino acid sequence alignment demonstrated that the amino acids in the putative activity...
switch (PAS) region of CsDFRa and CsDFRb1 are relatively conserved, with only four amino acids difference (Figure 7A). We conducted a series of gene mutation studies on different residues to further confirm that these distinct amino acids alter the DFR activity, and DHQ was used as a substrate to evaluate the reducing activity of the mutated proteins (Figure 7B). These results showed that the CsDFRa mutated proteins, CsDFRa$^{S117N}$ and CsDFRa$^{T123C}$, lost the DHQ reducing activity. Meanwhile, the DFR activity was restored by CsDFRb1 mutations, CsDFRb1$^{N120S}$ and CsDFRb1$^{C126T}$. The mutations in the other two different residues, such as DFRa$^{I118V}$ and DFRa$^{I119M}$, had no effect on the activity of the non-mutated protein (data not shown). The results of site-directed mutagenesis showed that the serine at 117 and threonine at 123 of CsDFRa are critical for modulating DFR reducing activity. Interestingly, these key amino acids are both belong to the hydroxyl amino acid family.

CsDFRa and CsDFRb1 homologous modeling was carried out using dihydroquercetin (DHQ) and NADPH as substrates, and splicing experiments were conducted using grape DFR crystals (PDB code 3BXX) $^{26}$. There was no significant difference in the binding of DFR residues to the substrate and NADPH in CsDFRs and site-directed mutated proteins (Figure 7B). Furthermore, the PAS region was found in helix α4, and the X-ray 3D structure revealed that this helix promotes water molecule binding to the carbonyl group of the asparagine main chain and the amino group of the catalytic lysine $^{27}$. This is the first of a cluster of water molecules that transfers a proton from the bulk solvent to the catalytic lysine $^{28,29}$. 


Figure 7. Identification of key amino acids that is critical for modulating DFR reducing activity. (A) Analysis of enzymes activity with different site-directed mutagenesis of the activity switch region. The red asterisk indicates the different amino acids of DFRα and DFRβ1 at the putative activity switch region, DHQ was used as a substrate to evaluate the reducing activity of the mutated proteins. (B) Homology model of DFRα, DFRβ1 and site-directed mutagenesis with docked dihydroquercetin (DHQ) and NADPH ligand, the PAS region was located in the helix α4 (red band). The model of homology was built using grape DFR crystals as a template (PDB code 3BXX).

Catalytic activities of recombinant CsDFR proteins in vitro

The dihydroflavonol substrate kinetic parameters of recombinant CsDFR proteins, including
CsDFRa, CsDFRc, CsDFRb\textsuperscript{N120S} and CsDFRb\textsuperscript{C126T}, were determined in potassium phosphate buffer at pH 7. CsDFRa $K_m$ values with DHK, DHQ, and DHM were 145.10, 41.80, and 58.44 μM, respectively. Meanwhile, for CsDFRc, the $K_m$ values for DHK, DHQ, and DHM were 42.31, 81.80, and 105.56 μM, respectively. The results indicated that CsDFRa may have a higher affinity for DHQ, whereas CsDFRc has a higher affinity for DHK (Figure 8 and Table S3).

Enzymatic assays were used to evaluate the activities of the site-directed mutated proteins of CsDFRb\textsuperscript{N120S} and CsDFRb\textsuperscript{C126T}. CsDFRb\textsuperscript{N120S} $K_m$ values for DHK, DHQ, and DHM were 148.96, 108.66, and 96.58 μM, respectively, while CsDFRb\textsuperscript{C126T} $K_m$ values were 131.09, 102.57, and 47.75 μM. These findings revealed that the N120S and C126T substitution for CsDFRb1 exhibit a higher affinity for the substrate DHM (Figure 8 and Table S3).
Figure 8. Kinetic parameters for DFRa, DFRc, DFRb1N120S and DFRb1C126T. In each enzyme reaction assay, two micrograms with purified recombinant proteins were used. A uniform concentration of 0.05 mg dihydroflavonol (DHK, DHQ, and DHM) was used as substrate. The data are expressed as means ± SD (n=3).

Discussion

CsDFR involvement in the biosynthesis of flavonoids in C. sinensis

DFR catalyzes the NADPH-dependent reduction of three different dihydroflavones to form the corresponding leucoanthocyanins. Leucoanthocyanins are the precursors for the synthesis of anthocyanins and flavan-3-ols. Therefore, DFR is the pivotal enzyme responsible for anthocyanins, catechins, and proanthocyanidins accumulation in plants. Catechins, procyanidins, and flavonols are the most abundant flavonoids in C. sinensis, with catechins mainly accumulated in buds and young leaves, while procyanidins are mainly accumulated in roots.

Numerous studies suggest that DFR expression is positively correlated with anthocyanin and PAs accumulation. In Saussurea medusa, for example, the SmDFR gene was expressed highly in flowers but significantly low in leaves and roots. In Lilium asiatica, two DFR genes were highly expressed in the colored parts of the perianth, anther, filament, stigma, and scale. Shimada investigated the tissue expression specificity of DFR using Bermuda fruit, flower, leaf, stem, and root as materials, and demonstrated that the tissue expression specificity of DFR gene was consistent with its catalytic activity. Tang LK demonstrated that AtDFR was highly expressed during pollen tube development. The results suggested that AtDFR may play an important role in pollen maturation and growth.

In tea plants, Mamati showed that the DFR gene was often highly expressed in tissues with high catechin content. Four DFR genes were identified and described from C. sinensis in this work. Among those genes, CsDFRa is strongly expressed in the bud and young leaf of C. sinensis, implying that CsDFRa is the major gene determining the accumulation of catechins in tea plant. While the other genes, on the other hand, are significantly expressed in the young root of tea plant, indicating that the three genes are linked to PAs buildup in the root.
Catalytic specificity analysis of the DFR family

Dihydroflavonol is classified into three types based on the number hydroxyls groups on the B-ring, they include DHK (mono-hydroxyl in B-ring), DHQ (di-hydroxyl in B-ring) and DHM (tri-hydroxyl in B-ring). In this study, enzyme activity analysis showed that four DFR enzyme exhibited different selectivities to three substrates in vitro. According to the structure analysis of proteins, DFR contains two important domains: the N-terminal highly conserved “NADPH” binding site and the substrate “dihydroflavones” binding site.

Previous studies has shown that the amino acid 134 is important for the catalytic activity of DFR protein. Therefore, based on the type of amino acid 134, DFR can be divided into three categories. Type 1, when position 134 is Aspartic acid (Asp), it is called the Asp-type, which can catalyze DHK to form the leucoanthocyanins. This type of DFR protein has been reported in cranberry. Type 2, when the position of amino acid 134 is asparagine (Asn), it is called the Asn-type, which can't catalyze the DHK to form leucoanthocyanins. Type 3, when position of amino acid 134 is neither Asp nor Asn, it is called non-Asp or Asn-type.

With the development of bioinformatics, DFRs have been identified in numerous plant species. Meanwhile, the previously identified genes of DFR family can also be divided into different types. For instance, in the Lotus corniculatus Linn, five DFR genes can be divided into three types. Among them, LcDFR1 belongs to the non-Asp or Asn-type, LcDFR2 and LcDFR3 belong to the Asp-type, the other two, LcDFR4 and LcDFR5, belong to the Asn-type.

In this study, we found that the understanding of the basic profile of DFR involvement in the catechin biosynthesis in C. sinensis is still being unraveled. The evolutionary analysis shows that CsDFRa belongs to Class I, and the function of most genes in this subfamily has been confirmed to have dihydroflavonol reductase activity (Figure 2B). Members of Class I are classified as Asn or Asp types based on the difference in the key amino acid sites in the substrate binding region. The kinetic analysis of recombinant enzyme revealed that CsDFRa had a high DHQ conversion efficiency.
The rooted phylogenetic tree analysis shows that CsDFRb1, CsDFRb3 are grouped into the Class II and CsDFRc are grouped into class III, which belongs to the non-Asp or Asn-type. The enzymatic activities of members of this subfamily have not been reported before. In the present study, we demonstrate that CsDFRc has the catalytic activity of dihydroflavonol 4-reductase by the functional analysis of prokaryotic recombinant protein and *Arabidopsis transgenic in vitro* and *in vivo*, respectively. The kinetic analysis of the recombinant enzyme showed that CsDFRc displayed highest conversion efficiency for DHK. However, CsDFRb1 and CsDFRb3 did not detect the catalytic activity of Dihydroflavonol 4-reductase.

**Crucial amino acid residues or regions for DFR protein**

Several studies have indicated that the DFR substrate specificity is determined by a 26 amino acid region. CsDFRb1 without catalytic activity and CsDFRa with catalytic activity were truncated and recombinated for each other, and the recombinant protein was expressed in *E. coli* to further investigate the key amino acid residues that determine the function of dihydroflavonol reductase. Analysis of enzyme activity showed that the twenty amino acid residues before the substrate binding region might have important effects on DFR enzyme activity. Furthermore, site-directed mutation confirmed that N120 and T126 are the key amino acid sites determining the catalytic activity of CsDFRb. Interestingly, among them, CsDFRb1\(^{N120S}\) and CsDFRb1\(^{T126C}\) have catalytic activity, and their affinity for DHM is significantly higher than that of DHQ and DHK. These results can be used for metabolic engineering of B-ring trihydroxyflavonoid.

Homologous modeling and substrate docking showed that the site-directed mutation had no effect on the three-dimensional structure of the DFR protein, and there was no significant difference in the spatial conformation between the substrate and NADPH. The putative activity switch (PAS) region is located in helix \(\alpha_4\) region (Figure 7B), which mainly determines the second \(\text{H}^+\) supply in the reduction reaction. Our findings confirmed that the key amino acid residues, serine and cysteine in PAS region (S117 and C123 of CsDFRa), are mercapto amino acid or hydroxyl amino acid, which may participate in the transfer of \(\text{H}^+\). The hydroxyl group established H-bond with water molecule or protein is critical for maintaining the reductase activity. The establishment of the H-bond with water molecule and proteins can’t be refuted, because the binding of a water
molecule to the carbonyl group of the main-chain of the asparagine (N120 of CsDFRb1) and the amino group of the catalytic lysine (T126 of CsDFRb1), they either break H-bond or induce steric hindrance for substrate binding, which coincides with previous findings. Taken together, these studies provide a theoretical foundation for learning more about DFR’s catalytic.

**Conclusion**

In conclusion, we isolated and characterized the CsDFR genes (CsDFRa, CsDFRc, CsDFRb1 and CsDFRb3) in *C. sinensis* and demonstrated their roles in *vitro* and *in vivo*. We found that its function is closely associated with catechins and PAs accumulation. Its physiological role was studied in the *Arabidopsis* tt3 mutant, and it was found that CsDFRa and CsDFRc recovered tt3 mutant phenotypes. Moreover, the enzyme activity assay was determined *in vitro* for the DFR activity of the CsDFRs and found that CsDFRa and CsDFRc have the DFR activity which converts dihydroflavonols into leucocyanidin. Through site-directed mutations, it was found that two amino acid residues of CsDFRs play an important role in enzyme activity. The enzymes kinetic revealed that the recombinant CsDFR enzyme showed that the optimal substrates for CsDFRa and CsDFRc were DHQ and DHK, and the optimal substrates for CsDFRb1N120S and CsDFRb1C126T were DHM, respectively. Thus, our data suggests that DFRs evolutionarily conserved in many plant species, which helps to provide a new candidate gene for the metabolic engineering of the flavonoids.

**Methods**

**Plant materials**

Different tissues bud, young leaf, mature leaf, stem and root of *C. sinensis* cv. Shuchazao were collected in early autumn from a five years old plant at Anhui Agricultural University in Hefei, China (north latitude 31.86, east longitude 117.27, altitude 20 m above mean sea level) experimental garden. The samples were frozen and stored at -80°C for future use.

The Landsberg erecta (Ler, wild-type) and *AtDFR* mutant (tt3, CS84) of Ler *Arabidopsis* were
obtained from the Arabidopsis Biological Resource Center (abrc.osu.edu) and grown in the growth chamber of the College of Life Sciences, Anhui Agricultural University with a 16 h light/8 h dark cycle at 22±2°C with a light intensity of 200 μmol m⁻² s⁻¹.

**Isolation and cloning of CsDFR genes**

RNAiso Plus for plant tissue kits (Takara, Dalian, China) was used to isolate total RNA from the bud and root. Standard end-to-end PCR reaction were performed on CsDFRs genes from the NCBI database, with gene specific primers designed (Table S2) according to the cDNA sequence. End-to-end PCR was carried out at 98°C for 30 s, 32 cycles at 98°C for 10 s, 58°C for 15 s, 72°C for 30 s, and a final extension at 72°C for 10 min. Subsequently, the PCR amplification products were purified using a Gel Extraction Kit (Aidlab, Beijing, China) and cloned into an Easy-Blunt vector (TransGen Biotech). The resulting vector was introduced into DH5α competent cells and sequenced.

**Phylogenetic analysis**

The unabridged length amino acid sequences of DFR proteins from *C. sinensis* and several other plants were acquired from the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for Phylogenetic analysis. ClustalW version 2.1 was used to align all of the sequences and deduced amino acid sequences of CsDFRs, and the alignment was then submitted to MEGA7 software (https://www.megasoftware.net) to generate a tree with neighbor-joining and bootstrapping (1000 replicates) analysis, as well as handling gaps with complete deletion.

**Quantitative real-time PCR analysis**

The gene-specific qRT-PCR primers were designed to study the tempo-spatial expression patterns of CsDFR genes in *C. sinensis* (Table S2). On a Peltier Thermal Cycler PTC200 (Bio-Rad, USA), the transcript level was determined using the SYBR premix Ex Taq II kit (Takara, Japan) Triplicate of all biological samples were analyzed. The Ct values of the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) from *C. sinensis* were used to normalize the reaction.
Heterologous expression of CsDFRs in *A. thaliana*

Using Gateway technology (Invitrogen, USA) the full-length *CsDFRs* cDNA was shuttled into the pCB2004 binary vector. Subsequently, the positive recombinant pCB2004-CsDFRs vector was *Agrobacterium tumefaciens* (GV3101) using electroporation, and the *Agrobacterium* was then used for floral dip transformation of *Arabidopsis* mutant *tt3* \(^3\). The transformed plants were grown in the glasshouse under carefully controlled conditions (16 h light/8 h dark cycle at 22±2°C with a relative humidity of 40%).

**Extraction and quantification analysis of anthocyanins and PAs from plant materials**

For the extraction of anthocyanins and PAs, briefly, freeze-dried samples (including seedlings and seeds) of 0.05 g of *CsDFRs* transgenic lines, WT and *tt3* mutant were ground by grinders. Subsequently, the total polyphenols were extracted at a low temperature with 1.2 mL of extraction solution A (HCl 0.5%: methanol 80%: water 19.5%, v/v) to the samples and vortexed and sonicated for 30 min at 4°C \(^7\). The samples were centrifuged at 13000g for 15 minutes before being re-extracted twice more as indicated above until the final volume of supernatants was 2mL. Finally, the combined solution was centrifuged for 10 minutes at 13 000g and the supernatants were stored at −20°C prior to total content analysis of anthocyanins and PAs. The contents of total anthocyanins in the supernatants were measured with UV spectrometer at 530 nm. After a reaction with the DMACA reagent (0.2% DMACA, w/v, HCl/methanol 1:7, v/v), the total soluble PAs content in the seeds was determined by a UV spectrometer at 640 nm. Briefly, the reaction mixture (1.3 mL total volume) contains 0.3 mL of supernatants and 1 mL of DMACA reagent. Subsequently, the reaction mixture was incubated at 25°C for 5 min, and the absorbent values of the reaction mixture were measured at 640 nm.

**Heterologous expression of CsDFRs in *Escherichia coli***

The CsDFRs ORF was subcloned into the pET-SUMO vector (Invitrogen, California, USA) in its entirety. The identity of the cloned gene was validated using T7 primers and sequencing analysis. The resulting vector SUMO-CsDFRs were transformed into BL21 (DE3) *E. coli* competent cells,
according to the manufacturer’s instructions. All of the chimeric and point mutation genes were cloned into the pET-SUMO vector with the same method as above. Positive colony, expression strains were cultivated in 200 mL of LB medium with 50 μg·mL⁻¹ kanamycin at 37°C. To promote recombinant protein expression, 1 mM isopropyl-β-D-thiogalactoside (IPTG) was added when the E.coil OD₆₀₀ reached 0.6. The cells were collected by centrifugation after 8 hours of induction at 28°C, and the fusion protein was purified using His-tag affinity chromatography according to the manufacturer’s instructions. Twelve percent SDS polyacrylamide gel staining and Coomassie brilliant blue staining were used to validate the isolated protein fraction. The activity assay in vitro was performed using pure recombinant proteins. All specific primers for SUMO-CsDFRs, chimeric DFRb1/DFRa, and amino acid point mutant genes were listed in Table S2.

**Enzyme activity assay and analysis of kinetics parameters**

The enzymatic reaction consisted of 0.1M pH 7.0 potassium phosphate buffer, 2 mM NADPH, 0.5 mM dihydroflavonols (DHK, DHQ and DHM) and 0.05 mg total soluble proteins content. The reaction mixtures were incubated at 30°C for 30 min, and the total volume was 200 μl. Because leucoanthocyanidin is unstable in solution and difficult to detect directly using HPLC, a three times volume of butanol–HCl reagent (95:5, v/v) was added, and the reaction was splitted at 95°C for 1h to form anthocyanidin and was centrifuged at 4000g for 15min. The absorption of the supernatants was measured at OD₅₃₀ nm by UV spectrophotometer. According to the absorption value of anthocyanin standard at different concentrations, the standard curve was established to quantify the content of products in the sample. To determine the apparent $K_m$ value of the reaction mixture of different substrates, 0.05 mg purified recombinant enzyme and three dihydroflavonols at a series of concentrations. The enzymatic products were quantified as the standard curve of anthocyanins standard. Apparent $K_m$ data were calculated using Hanes plots.

**Molecular Docking statistic and site-directed mutagenesis**

The homology model of CsDFRa and CsDFRb1 were constructed with the crystal structure of grape DFR (PDB code 3BXX) as a template using the online server SWISS-MODEL.
Furthermore, dihydroquercetin and NADPH were taken as the substrates, and they were docked into the model of CsDFRa and CsDFRb1 with the lowest energy conformation, using the software AutodockTools-1.5.6 (http://mgltools.scripps.edu/downloads). The resulting spatial model structure of the CsDFRa and CsDFRb1 protein, including binding sites and substrate linkage pockets, were analyzed and visualized using Python 27 software. The site-directed mutation plasmids CsDFRa<sub>S117N</sub>, CsDFRa<sub>T123C</sub>, CsDFRb1<sub>N120S</sub>, CsDFRb1<sub>C127T</sub>, DFRa<sub>I178V</sub> and DFRa<sub>I179M</sub> were generated using an overlap extension PCR technology with CsDFRa and CsDFRb1 plasmids as the templates, respectively. All of the specific oligonucleotide primers used to construct the mutations plasmids is listed in Table S2.

**Statistical methods**

Three separate experiments or three biological replicates were used in each of the aforementioned test. Using SPSS software, mean values were generated and then statistically analyzed using a Student’s test. The two-tailed test of significance was used. Statistical significance was determined by P-values of less than 0.05.

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China (31770729, 31870676); the National Key R&D Program of China (2018YFD1000601); Collegiate Collaborative Innovation Foundation of Anhui Province (GXXT-2020-081) and Joint Funds of the National Natural Science Foundation of China (U21A20232).

**Author contributions**

RHX, SXX and WYS conceived and designed the research. WYS, GLP and XT conducted experiments. SXX, RHX, LY, LT and DXL performed the experiments. RHX, RA and WYS analyzed the data. RHX and DXL wrote the manuscript, RA and WYS edited and contributed to the content.

**Conflict of Interest**

There are no conflicts of interest declared by any of the writers.
Data availability statement

All data supporting the conclusions of this study may be found in the publication and its supplemental materials, which are available online. Any additional relevant information can be obtained from the corresponding author upon request (Yunsheng Wang).

Supporting information

Figure S1: Schematic representation of CsDFRs gene structures and promoter elements.

Table S1: Relative quantity of different flavonoid compounds in different tissues.

Table S2: List primers used in this study.

Table S3: Kinetic parameters for DFRa, DFRc, DFRb1^{N120S} and DFRb1^{C126T}.

References

1. Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol*. 2001;126:485-93.

2. Cabrera C, Artacho R, Gimenez R. Beneficial effects of green tea--a review. *J Am Coll Nutr*. 2006;25:79-99.

3. Chen A Y, Chen Y C. A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. *Food Chem*. 2013;138:2099-107.

4. Dalgaard F, Bondonno N P, Murray K et al. Associations between habitual flavonoid intake and hospital admissions for atherosclerotic cardiovascular disease: a prospective cohort study. *Lancet Planet Health*. 2019;3:e450-e59.

5. Tanaka Y, Brugliera F. Flower colour and cytochromes P450. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 2013;368:20120432.

6. Hua C, Linling L, Shuiyuan C et al. Molecular cloning and characterization of three genes encoding dihydroflavonol-4-reductase from Ginkgo biloba in anthocyanin biosynthetic pathway. *PLoS One*. 2013;8:e72017.
7. Jiang X, Liu Y, Li W et al. Tissue-specific, development-dependent phenolic compounds accumulation profile and gene expression pattern in tea plant [Camellia sinensis]. *PLoS One.* 2013;8:e62315.

8. Bowers J E, Dangl G S, Vignani R et al. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (Vitis vinifera L.). *Genome.* 1996;39:628-33.

9. Helariutta Y, Elomaa P, Kotilainen M et al. Cloning of cDNA coding for dihydroflavonol-4-reductase (DFR) and characterization of dfr expression in the corollas of Gerbera hybrida var. Regina (Compositae). *Plant Mol Biol.* 1993;22:183-93.

10. Wang J, Wang Y, Wang S et al. Cloning and temporal-spatial expression analysis of dfr gene from Scutellaria baicalensis with different colors. *Sheng Wu Gong Cheng Xue Bao.* 2021;37:1312-23.

11. Wang L, Zhu Y, Wang P et al. Functional Characterization of a Dihydroflavanol 4-Reductase from the Fiber of Upland Cotton (Gossypium hirsutum). *Molecules.* 2016;21:32.

12. Bashandy H, Pletiainen M, Carvalho E et al. Anthocyanin biosynthesis in gerbera cultivar 'Estelle' and its acyanic sport 'Ivory'. *Planta.* 2015;242:601-11.

13. Kubasek W L, Shirley B W, Mckillop A et al. Regulation of Flavonoid Biosynthetic Genes in Germinating Arabidopsis Seedlings. *Plant Cell.* 1992;4:1229-36.
14. Johnson E T, Yi H, Shin B et al. Cymbidium hybrida dihydroflavonol 4-reductase does not efficiently reduce dihydrokaempferol to produce orange pelargonidin-type anthocyanins. *Plant J.* 1999; **19**:81-5.

15. Johnson E T, Ryu S, Yi H et al. Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. *Plant J.* 2001; **25**:325-33.

16. Alagawany M, Abd El-Hack M E, Saeed M et al. Nutritional applications and beneficial health applications of green tea and L-theanine in some animal species: A review. *J Anim Physiol Anim Nutr (Berl).* 2020; **104**:245-56.

17. Yao S S, Guo W F, Lu Y et al. Flavor characteristics of lapsang souchong and smoked lapsang souchong, a special Chinese black tea with pine smoking process. *J Agric Food Chem.* 2005; **53**:8688-93.

18. Asakawa T, Hamashima Y, Kan T. Chemical synthesis of tea polyphenols and related compounds. *Curr Pharm Des.* 2013; **19**:6207-17.

19. Hu J, Zhou D, Chen Y. Preparation and antioxidant activity of green tea extract enriched in epigallocatechin (EGC) and epigallocatechin gallate (EGCG). *J Agric Food Chem.* 2009; **57**:1349-53.

20. Rani A, Singh K, Ahuja P S et al. Molecular regulation of catechins biosynthesis in tea [Camellia sinensis (L.) O. Kuntze]. *Gene.* 2012; **495**:205-10.

21. Wang W, Zhou Y, Wu Y et al. Insight into Catechins Metabolic Pathways of Camellia sinensis Based on Genome and Transcriptome Analysis. *J Agric Food Chem.* 2018; **66**:4281-93.
22. Katsu K, Suzuki R, Tsuchiya W et al. A new buckwheat dihydroflavonol 4-reductase (DFR), with a unique substrate binding structure, has altered substrate specificity. *Bmc Plant Biol.* 2017;17:239.

23. Shirley B W, Hanley S, Goodman H M. Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis transparent testa mutations. *Plant Cell.* 1992;4:333-47.

24. Pang Y, Abeysinghe I S, He J et al. Functional characterization of proanthocyanidin pathway enzymes from tea and their application for metabolic engineering. *Plant Physiol.* 2013;161:1103-16.

25. Xie D Y, Sharma S B, Paiva N L et al. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science.* 2003;299:396-9.

26. Trabelsi N, Petit P, Manigand C et al. Structural evidence for the inhibition of grape dihydroflavonol 4-reductase by flavonols. *Acta Crystallogr D Biol Crystallogr.* 2008;D64:883-91.

27. Petit P, Granier T, D'estaintot B L et al. Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis. *J Mol Biol.* 2007;368:1345-57.

28. Benach J, Filling C, Oppermann U C et al. Structure of bacterial 3beta/17beta-hydroxysteroid dehydrogenase at 1.2 A resolution: a model for multiple steroid recognition. *Biochemistry.* 2002;41:14659-68.

29. Bottoms C A, Smith P E, Tanner J J. A structurally conserved water molecule in Rossmann dinucleotide-binding domains. *Protein science : a publication of the Protein Society.* 2002;11:2125-37.
30. Zhao J, Li P, Xia T et al. Exploring plant metabolic genomics: chemical diversity, metabolic complexity in the biosynthesis and transport of specialized metabolites with the tea plant as a model. Crit Rev Biotechnol. 2020;40:667-88.

31. Li H, Qiu J, Chen F et al. Molecular characterization and expression analysis of dihydroflavonol 4-reductase (DFR) gene in Saussurea medusa. Mol Biol Rep. 2012;39:2991-9.

32. Nakatsuka A, Izumi Y, Yamagishi M. Spatial and temporal expression of chalcone synthase and dihydroflavonol 4-reductase genes in the Asiatic hybrid lily. Plant Science. 2003;165:759-67.

33. Shimada N, Sasaki R, Sato S et al. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the Lotus japonicus genome. Journal Of Experimental Botany. 2005;56:2573-85.

34. Tang L K, Chu H, Yip W K et al. An anther-specific dihydroflavonol 4-reductase-like gene (DRL1) is essential for male fertility in Arabidopsis. New Phytol. 2009;181:576-87.

35. Mamati G E, Liang Y R, Lu J L. Expression of basic genes involved in tea polyphenol synthesis in relation to accumulation of catechins and total tea polyphenols. J Sci Food Agr. 2006;86:459-64.

36. Martens S, Teeri T, Forkmann G. Heterologous expression of dihydroflavonol 4-reductases from various plants. FEBS Lett. 2002;531:453-8.
37. Polashock J J, Griesbach R J, Sullivan R F et al. Cloning of a cDNA encoding the cranberry dihydroflavonol-4-reductase (DFR) and expression in transgenic tobacco. *Plant Science*. 2002;163:241-51.

38. Shimada N, Sasaki R, Sato S et al. A comprehensive analysis of six dihydroflavonol 4-reductases encoded by a gene cluster of the Lotus japonicus genome. *J Exp Bot*. 2005;56:2573-85.

39. Clough S J, Bent A F. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J*. 1998;16:735-43.