Cloning and developmental expression analysis of a basic helix loop helix transcription factor gene for epigallocatechin-3-gallate biosynthesis in tea plant (Camellia sinensis L.)

Shizhong Zheng¹²³, Yuling Lin¹³, Ping Sun¹³, Zhongxiong Lai¹³*, Jinke Lin¹* ¹College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou, 350002, China, ²Ningde Normal University, Ningde 352100, China; ³Institute of Horticultural Biotechnology, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

Received 1 April 2015; revised 9 July 2015; accepted 20 July 2015

Epigallocatechin-3-gallate (EGCG) is the most effective active tea polyphenol, with a wide usage in food and medicine. Basic helix-loop-helix (bHLH) transcription factors play an important role in the regulation of flavonoid biosynthesis, but it remains unclear how the bHLH gene regulates the EGCG biosynthesis in the tea plant. We isolated and cloned the full-length bHLH gene from Camellia sinensis (CS_bHLH), which is 1,599 bp in length, and encodes a 379 amino acid protein of predicted molecular weight of 41.69 kDa. Bioinformatics analysis showed that the encoded protein is acidic, hydrophilic, unstable, and without the signal peptide. Its secondary structure mainly consists of random coils (70.2%) and α-helixes (22.4%). It shares 68.97% protein sequence identity with bHLH from other plants. The EGCG content of '1005' exceeded that of its parent cultivars in the shoot and second leaf, and that's up to 12.48% in shoot of '1005'. Comparing the expression level of CS_bHLH and the EGCG content at different developmental stages of three cultivars, it was found a similar trend for each cultivar: bud > second leaf > fourth leaf. For the samples of each cultivar, the higher the EGCG content, the higher the CS_bHLH expression level. That revealed that CS_bHLH may positively regulate the biosynthesis of EGCG.

Keywords: bHLH transcription factor, bioinformatics, developmental stages, EGCG, tea plant (Camellia sinensis)

Introduction

Thousands of years ago, the leaf and bud of Camellia sinensis has been used for tea beverage. A lot of studies suggest that tea has diverse biological activities since it contains valuable secondary metabolites, such as tea polyphenols, theanine. Epigallocatechin-3-gallate (EGCG) is the most effective component of tea polyphenols. As a type of flavonoid, EGCG is reported to have many health benefits, including anti-oxidation, anti-cancer, anti-radiation, anti-diabetic, anti-hyperlipidemia. Indeed, the drinking of tea, especially that rich in EGCG, is thought to improve the immune system, lower blood lipid levels, prevent diabetes, and have anti-aging and anti-cancer effects. Additionally, it is widely used in food processing, where it demonstrates anti-oxidation activity that is stronger than that of vitamin E.

The EGCG content varies according to the developmental stage of the leaves of the tea plant (Camellia sinensis L.). Although the basic biosynthetic pathway of flavonoids has been reported and related components described, the molecular mechanism of flavonoid biosynthesis in the tea plant remains unclear. Similarly, the regulation of genes in the EGCG biosynthesis pathway of the tea plant is poorly understood, even though a number of structural genes of flavonoid biosynthesis have been cloned or isolated, including phenylalanine ammonia lyase, anthocyanidin reductase, anthocyanidin synthase, dihydroflavonol-4-reductase, chalcone synthase, chalcone isomerase, flavanone 3-hydroxylase, and leucoanthocyanidin reductase. Recent studies suggest that a ternary complex MYB-bHLH-WD40 (MBW) orchestrates the regulation of structural gene expression in the flavonoid pathway.

Recent studies about the regulation of transcription factors in the flavonoid pathway mostly focused on MYB, but reports of the function of bHLH proteins in the flavonoid pathway are fewer.

Basic helix-loop-helix (bHLH) proteins are a family of transcription factors that are widespread in
eukaryotes, with 173 and 158 in *Oryza sativa* and *Arabidopsis* thaliana, respectively\(^\text{18}\). They play an important role in secondary metabolism, signal transduction, stress response, and organogenesis\(^\text{19-23}\). Based on our preliminary work, we hypothesized that bHLH might be involved in the regulation of EGCG biosynthesis in the tea plant *C. sinensis*\(^\text{24}\). The *C. sinensis* germplasm used in the present study, known as ‘1005’, is rich in EGCG and has male and female parent plants that are cultivars of ‘Huangdan’ and ‘Fuyun 7’, respectively.

We isolated and cloned the full-length bHLH gene (*CS_bHLH*) from the ‘1005’ tea plant using rapid amplification of cDNA ends (RACE) technology. Using shoots and leaves at different developmental stages of three cultivars (strains), we analyzed the expression level of *CS_bHLH* and determined the EGCG content to help elucidate the function of *CS_bHLH* in EGCG biosynthesis.

**Materials and Methods**

**Chemicals and Reagents**

The EGCG standard was purchased from TCI Development Co., Ltd (Shanghai, China), and the purity is no less than 98%. Acetonitrile-methanol and acetic acid which are HPLC grade were also purchased from TCI Development Co., Ltd. All the other chemicals and solvents for HPLC were high analytical grade. Double distilled water was purified from Milli-Q water system. The stable working solution was prepared with EDTA (10 mg/mL)/ascorbic acid (10 mg/mL)/acetonitrile/double distilled water (25/25/50/400; V/V/V/V).

**Plant Samples**

Samples were collected from the ‘1005’ cultivar (offspring), ‘Huangdan’ (male parent) and ‘Fuyun 7’ (female parent) growing in an experimental field at Fujian Agriculture and Forestry University. We collected fresh shoots, second leaves and fourth leaves from the same plant. Every sample was divided into two parts. The first part was dried for EGCG extraction and determination, and the other one was immediately frozen in liquid nitrogen and stored at -80°C until required.

**RNA Extraction, cDNA Synthesis and CS_bHLH Cloning**

Total RNA was extracted from the shoots of the ‘1005’ plant using the TransZolTM UP Plus RNA Kit (TransGen Biotech, Beijing, China). First-strand cDNA was then synthesized with the SMARTerTM RACE cDNA Amplification Kit (TAKARA, Beijing, China) and Revert Aid First Strand cDNA Synthesis Kit (Thermo Scientific, Shanghai, China). We used RACE to obtain the full-length *CS_bHLH* with primers designed using DNAMAN version 6.0 software and based on the *CS_bHLH* expressed sequence tag (Table 1). Primers were synthesized by BGI Tech (Shenzhen, China). Thermocycling parameters were: 95°C for 3 min, then 35 cycles of 94°C for 30 s, the primer melting temperature for 30 s, and 72°C for 2 min, followed by a final elongation at 72°C for 10 min. The amplified product was purified from a 1% agarose gel using the Gene JET Gel Extraction Kit (Thermo Scientific) and ligated to the pMD18-T Vector (TAKARA) at 16°C. The ligation product was used to transform DH5α competent cells and the transformants were selected on ampicillin plates. Single resistant colonies were screened by PCR and positive ones sequenced by BGI Tech. Sequence data were compared against published bHLH sequences from other plants using the Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi).

<table>
<thead>
<tr>
<th>Primer(5’-3’)</th>
<th>Tm</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ORF Cloning</strong></td>
<td></td>
</tr>
<tr>
<td>Forward: 5’-GCGGAGAAGATCAAGGAATTC-3’</td>
<td>53.5°C</td>
</tr>
<tr>
<td>Reverse: 5’-TTGGATAGACACCCTGCATTG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>5’RACE</strong></td>
<td>58.5°C</td>
</tr>
<tr>
<td>Outer primer: 5’-ATACTACAGTCTGAACCCGAGC-3’</td>
<td></td>
</tr>
<tr>
<td>UPM: 5’-CTTATACGACACTCATATAGGGCAACAGTGGTATCAACGCAGT-3’</td>
<td></td>
</tr>
<tr>
<td>Inner primer: 5’-GAGAAGGAGACACATATTAGTAG-3’</td>
<td>56.5°C</td>
</tr>
<tr>
<td>NUP: 5’-AAGCAGTGGTATCAACGCGAGT-3’</td>
<td></td>
</tr>
<tr>
<td><strong>3’RACE</strong></td>
<td>56.5°C</td>
</tr>
<tr>
<td>Out primer: 5’-CGGCACATAGGATTCGTACAAC-3’</td>
<td></td>
</tr>
<tr>
<td>Inner primer: 5’-TCTGATACGCGACGTTCCTCCTC-3’</td>
<td>57°C</td>
</tr>
<tr>
<td>AUP: 5’-GGCCACGCGTCTAGACTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3’</td>
<td></td>
</tr>
</tbody>
</table>
Bioinformatic Analysis of CS_bHLH

The full-length CS_bHLH sequence was analyzed using the following: DNAMAN version 6.0 software and ORF Finder (http://www.ncbi.nlm.nih.gov/projects/orffinder/) to predict the open reading frame and amino acid sequences, and sequence alignment of homologous proteins; the ExPASy ProtParam to predict and analyze physicochemical properties (http://web.expasy.org/protparam/); NCBI online software to determine conserved domains (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi); PSORT for subcellular localization prediction (http://psort.hgc.jp/form.html); SignalP 3.0 Server for signal peptide prediction (http://www.cbs.dtu.dk/services/SignalP-3.0/); Protscal for hydrophobic prediction (http://www.expasy.org/cgi-bin/protscal/protscal.pl); NetPhos2.0 Server for protein phosphorylation site prediction (http://www.cbs.dtu.dk/services/NetPhos/); PSIPRED for secondary and tertiary protein structure prediction (http://bioinf.cs.ucl.ac.uk/psipred/); SWISS-MODEL for protein structure homology modelling (http://swissmodel.expasy.org/). We also used Molecular Evolutionary Genetic Analysis version 5.0 software to conduct phylogenetics analysis\textsuperscript{25}, and constructed a phylogenetic tree using a neighbor-joining algorithm. Bootstrap values were computed with 1,000 replicates to evaluate support for the putative groupings.

Real-time Reverse Transcriptase PCR Analysis of CS_bHLH

Quantitative real-time reverse transcriptase (qRT)-PCR was performed using the LightCycler 480\textsuperscript{®} Real-Time PCR system (Roche). Total RNA from leaves at different developmental stages was extracted as described above. Reverse transcription was carried out using the PrimeScript\textsuperscript{TM} RT reagent kit (TAKARA) on 1 \( \mu \)g of total RNA. The forward primer (5′-GGAGAGTCAAACCAGCAACAC-3′) and reverse primer (5′-AGCAGGGGAAAAGA GAAGGAG-3′) were designed and synthesized as described above, and amplified a product 111 bp in length. qRT-PCR was carried out as follows: 95°C for 30 s, then 45 cycles of 95°C for 5 s, 59°C for 30 s, followed by analysis of the melting curves at 60-95°C. The glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was amplified as an internal reference using primers 5′-TTGGCATTGTTAGGGTCTC-3′ (forward) and 5′-CAGTGGGAACCGGAAAGC-3′ (reverse). Data were analyzed using \( Q = E^{-\Delta\Delta C_T} \) to confirm the relative expression of the various samples\textsuperscript{36}.

Analysis of EGCG Content in Developmental Leaves

For calibration curve, mother solution of standard EGCG of 2.00 mg/mL was prepared firstly, then it was diluted in the stable working solution. The standard solutions were arranged in the 0.50-3.00 \( \mu \)g range. Of every standard solutions 10 \( \mu \)L was injected. Calibration curves (\( Y=3\times10^{-7}X-0.0010, R^2=0.999; \ Y: \text{Peak area}; \ X: \text{Injection quantity} \) was done using standard solutions at 6 different concentrations, between 50.00 \( \mu \)g/mL-300.00 \( \mu \)g/mL.

Each tea sample (5.0 g) was accurately weighed (\( \pm \) 0.0001). After being dried for 1h at 120°C, Each sample was accurately weighed (\( \pm \) 0.0001) right now. Then the dry weight of each sample could be calculated exactly. Each tea sample was milled (0.84 mm) of every milled sample, 0.2 g (\( \pm \) 0.0001) was accurately weighted and transferred into a 10 mL centrifuge tube. The tea sample was extracted with 5 mL of 70% hot methanol (70°C) at the temperature of 70°C for 10 min with intermittent shaking. The extracts were cooled to room temperature and centrifuged for 10 min (3500 r/min), then the supernatant fluid was transferred into a 10 mL volumetric flask. The operations were repeated twice. Total extracts were combined in the same volumetric flask, which was filled up to the calibration mark with 70% methanol. After being shaken up, the extracts were then filtered through 0.45 \( \mu \)m Millipore filter. Of each sample 10 \( \mu \)L was injected.

A high-performance liquid chromatography (HPLC) system (Thermo EASY-nLC 1000) was used for quantitative analysis. All the operations and data analysis were performed by the Chemstation software. An Agilent C18 column (4.6 \times 250 mm, 5-Micron) was used at 35°C. All solvents were filtered through 0.45 \( \mu \)m Millipore filter and degassed by stripping of helium. The gradient elution was carried out using the following solvent systems: mobile phase A, acetonitrile/acetic acid/EDTA (10 mg/mL)/double distilled water (90/20/2/888;V/V/V/V); mobile phase B, acetonitrile/acetic acid/EDTA (10 mg/mL)/double distilled water (800/20/2/178;V/V/V/V). All solvents were filtered through 0.45 \( \mu \)m Millipore filter. The linear gradient elution system was: 0-10 min, 100%; A: 10-15 min, from 100% A to 68% A and 32% B; standing at 68% A and 32% B for 10 min and returning to 100% A, after other 5 min. The flow rate was 1.0 mL/min and the quantification was performed at 278 nm.
Results

Cloning and Characterization of CS_bHLH

Full-length CS_bHLH was cloned from the ‘1005’ strain of C. sinensis using RACE. It was 1,599 bp long with an ORF of 1,140 bp that encoded a protein composed of 379 amino acids (Fig. 1). The 1,599 bp sequence has been deposited into GenBank under accession number KM236565. Bioinformatic analysis of CS_bHLH showed that it had a molecular formula of C_{2586}H_{4254}N_{740}O_{775}S_{21} with a predicted molecular weight of 41.69 kDa. The most common amino acid was alanine (10.3%), and cysteine was the least common (1.1%). The protein was acidic with pI 4.54. It was also predicted to be unstable and hydrophilic without its signal peptide. The CS_bHLH protein contained a conserved bHLH-MYC_N superfamily domain between the ninth and 191st amino acid residue, which we predict may be important for regulating EGCG.

PSORT predicted that the protein was most likely to be located in the mitochondrial matrix space and cytoplasm of the plant cell, with a score of 0.485 and 0.450, respectively. ProtScale predicted there to be fewer hydrophobic than hydrophilic amino acids, which was in agreement with the prediction of its hydrophilic nature. NetPhos2.0 predicted there to be 26 phosphorylated sites in the CS_bHLH protein, which are likely to play an important role in the regulation of transcription and metabolism, as well as the ability of the protein to bind DNA\(^2\). PSIPRED and SWISS-MODEL predicted CS_bHLH secondary (Fig. 2) and tertiary structures (Fig. 3), suggesting a mainly random coil (70.2%) and α-helical (22.4%) structure, with some β-strands (7.4%).

Homological and Phylogenetic Analysis of CS_bHLH

NCBI BLAST detected amino acid sequence homology of CS_bHLH with bHLH from other species, including Diospyros kaki (AEC03343; 78%), Vitis vinifera (NP_001268182; 80%), Ricinus communis (XP_002520758;75%), and Morella rubra.

Fig. 1 — CS_bHLH cDNA and deduced amino acid sequence (Initiation and termination codons are underlined).

Fig. 2 — Predicted secondary structure of CS_bHLH protein.
The CS_bHLH protein was found to share 68.97% identity with bHLH proteins from these same species (Fig. 4). Moreover, phylogenetic analysis revealed that CS_bHLH was most closely related to bHLH of D. kaki (Fig. 5).

The Relative Expression Levels of CS_bHLH and the EGCG Contents in Leaves at Different Development Stages

After being normalized against reference gene GAPDH for each sample, CS_bHLH expression was observed to vary according to the developmental stage of the leaves collected from the three cultivars (Fig. 6), with a similar trend observed for each cultivar: bud > second leaf > fourth leaf. Whereas
changes were most obvious for ‘1005’, and relative expression was about 4 times higher in shoot than that in 4th leaf. We determined the EGCG contents of all samples via HPLC (Fig. 7), this was also found to follow a trend similar to that of CS_bHLH expression for each cultivar. It was also found the EGCG content of ‘1005’ exceeded that of its parent cultivars in the shoot and second leaf. EGCG content with 12.48% in shoot of ‘1005’ was the highest, whereas that in 4th leaf of Huangdan was only 2.24%, the lowest.

Discussion

Some studies suggest that MYB doesn’t only regulate the flavonoid biosynthesis by themselves, but interacts with bHLH protein^{17,28}. The plant protein with bHLH domain which is involved in the regulation of flavonoid biosynthesis was firstly found in maize^{29}. Recent studies also suggest that bHLH is thought to be involved in the control of plant growth, including the biosynthesis of plant flavonoids^{30,31}. However, few studies have specifically investigated its function in the tea plant, particularly with regard to EGCG biosynthesis. In the present study, we firstly cloned the full-length bHLH gene from C. sinensis (CS_bHLH), and found that it appears to be closely associated with the biosynthesis of EGCG. Every member of the bHLH transcription factor family shares a conserved domain, and this falls between the ninth and 191st amino acid residue of CS_bHLH. These domains are very important for DNA binding^{32}. Phylogenetic analysis of CS_bHLH showed that it was clustered within the same clade as the bHLH from D. kaki named DKMYC1 (AEC03343), suggesting that they might share a similar function in regulating flavonoid biosynthesis. DKMYC1 was reported that it was involved in the proanthocyanidin and anthocyanin biosynthesis in persimmon^{33}.

Our analysis showed that EGCG content patterns were similar to those of CS_bHLH expression in the three tea plant cultivars. For the samples of each cultivar, the higher the EGCG content, the higher the CS_bHLH expression level, indicating that CS_bHLH might positively regulate the biosynthesis of EGCG. Although the mechanism by which this might occur requires further study, the most common regulation mechanisms of bHLH proteins include the formatting of double dimers and interaction with MYB transcription factors^{34,35}. Additionally, the bHLH MYC2 is thought to control the expression of MYB75/PAP1 and EGL3, which are positive regulators of flavonoid synthesis^{36}.

In conclusion, CS_bHLH was a new bHLH transcription factor isolated from tea plant. And we found that CS_bHLH expression appears to be associated with the EGCG content in the present study, suggesting that it might be involved in its synthesis by regulating the related structural gene. Though it still needs much effort for elucidating EGCG biosynthesis at the molecular level. These findings lay the groundwork for the further study of the biosynthesis of EGCG.

Acknowledgements

This work was supported financially by the National Natural Science Foundation of China (31170651). We are grateful to Dr. Guoyin Kai for his help in revising English composition.

References


