# Postprint: Cloning and Expression Analysis of the FtF5H Gene in Tartary Buckwheat

**Authors:** Duan Ying, Yang Xiaolin, Cai Suyun, He Runli, Yin Guifang, Wang Yanqing, Lu Wenjie, Sun Daowang, Wang Lihua

**Date:** 2021-07-20T00:00:00+00:00

## Abstract

Ferulate 5-hydroxylase (F5H) is a key enzyme regulating S-type lignin synthesis. To investigate its molecular mechanism in the lignin biosynthesis pathway of tartary buckwheat, this study screened and identified an F5H gene from tartary buckwheat transcriptome data, named FtF5H (GenBank accession number: MW455111). Bioinformatics methods were employed to analyze and predict the physicochemical properties, signal peptide, transmembrane structure, subcellular localization, hydrophilicity/hydrophobicity, protein secondary structure, protein tertiary structure, amino acid composition, and phylogenetic tree of the tartary buckwheat F5H protein. Simultaneously, real-time quantitative PCR (qRT-PCR) technology was used to analyze the differential expression of the FtF5H gene in leaves, flowers, stems, and hulls of thick-hulled and thin-hulled tartary buckwheat. The results showed that the FtF5H gene sequence contains a complete cDNA open reading frame of 1,395 bp, encoding 464 amino acids; bioinformatics prediction revealed that the FtF5H protein possesses a P450 superfamily structure, is a hydrophilic, stable, acidic protein, lacks transmembrane domains, and is a non-secretory protein; the secondary structure of the FtF5H protein is mainly composed of $\alpha$-helices and random coils; tertiary structure prediction showed that the FtF5H protein has high similarity with 5ylw.1.A. Phylogenetic analysis indicated that FtF5H belongs to the CYP84A subfamily. qRT-PCR revealed that the FtF5H gene is expressed in different tissues of both tartary buckwheat varieties, with expression in the hulls of thick-hulled tartary buckwheat being 5 times higher than that in thin-hulled tartary buckwheat, exhibiting a highly significant difference. This study lays a foundation for further investigation into the molecular regulatory mechanism of lignin synthesis in tartary buckwheat and is of significant importance for the breeding of new tartary buckwheat varieties.

# Full Text

## Preamble

### Cloning and Expression Analysis of the FtF5H Gene from Tartary Buckwheat

**DUAN Ying[1], YANG Xiaolin[1], CAI Suyun[1], HE Runli[1]*, YIN Guifang[2], WANG Yanqing[2], LU Wenjie[2], SUN Daowang[2], WANG Lihua[2]**

[1]College of Traditional Chinese Medicine and Food Engineering, Shanxi University of Chinese Medicine, Taiyuan 030619, China
[2]Biotechnology and Genetic Germplasm Resources Research Institute, Yunnan Academy of Agricultural Sciences/Yunnan Provincial Key Laboratory of Agricultural Biotechnology/Key Laboratory of Southwestern Crop Gene Resources and Germplasm Innovation, Ministry of Agriculture and Rural Affairs, Kunming 650201, China

## Abstract

Ferulate 5-hydroxylase (F5H) is a key enzyme regulating S-type lignin synthesis. To investigate its molecular mechanism in lignin biosynthesis pathways of tartary buckwheat, this study identified an F5H gene from tartary buckwheat transcriptome data, designated as FtF5H (GenBank accession number: MW455111). Bioinformatics methods were employed to analyze and predict the physicochemical properties, signal peptide, transmembrane structure, subcellular localization, hydrophilicity/hydrophobicity, secondary structure, tertiary structure, amino acid composition, and phylogenetic relationships of the FtF5H protein. Additionally, real-time quantitative PCR was used to analyze differential expression of FtF5H in leaves, flowers, stems, and husks of thick-husk and thin-husk tartary buckwheat varieties.

The results revealed that the FtF5H gene contains a complete cDNA open reading frame of 1,395 bp, encoding 464 amino acids. Bioinformatics prediction indicated that the FtF5H protein possesses a P450 superfamily structure, is a hydrophilic, stable, acidic protein lacking transmembrane domains, and is non-secretory. The secondary structure of FtF5H protein consists primarily of $\alpha$-helices and random coils. Tertiary structure modeling showed high similarity with 5ylw.1.A. Phylogenetic analysis demonstrated that FtF5H belongs to the CYP84A subfamily. qRT-PCR analysis revealed that FtF5H was expressed in all tested tissues of both buckwheat varieties, with expression in thick-husk buckwheat husks being five-fold higher than in thin-husk varieties, representing an extremely significant difference. This study establishes a foundation for further investigation into the molecular regulatory mechanisms of lignin synthesis in tartary buckwheat and holds important significance for breeding new tartary buckwheat varieties.

**Keywords:** tartary buckwheat; RT-PCR cloning; ferulate 5-hydroxylase; bioin-

formatics analysis; gene expression

## Introduction

Tartary buckwheat (*Fagopyrum tataricum*) is a dual-purpose crop for both medicinal and food uses, with bitter, neutral, and cold properties. It is traditionally believed to benefit energy, spirit, hearing, vision, and gastrointestinal function, earning it the reputation as the "king of five grains" and a "three-reduction" food (Zhu and Guo, 2014). Cultivated tartary buckwheat typically possesses thick husks with a husk ratio of 20%-30%. These tough, thick husks result in low dehulling rates ranging from 2% to 6% (Song et al., 2019), making it difficult to produce whole tartary buckwheat kernels and obtain high-rutin-content bran-layer nutrients. This limitation reduces the bioactive compounds and nutritional efficacy of tartary buckwheat (Chen et al., 2015). Additionally, during direct milling of tartary buckwheat seeds, small amounts of husk powder 混入 can decrease palatability and restrict deep processing of tartary buckwheat products, severely impacting industrial development.

'Rice buckwheat' is a local variety from Yunnan and Guizhou provinces that is easily dehulled, with dehulling rates reaching up to 93% (Song et al., 2019). As a special resource, 'rice buckwheat' holds important theoretical and practical significance for studying the genetic patterns, functional genes, variety breeding, and processing of thin-husk characteristics in tartary buckwheat.

F5H belongs to the cytochrome P450 monooxygenase family (Meyer et al., 1998). First isolated from sweetgum in 1996, it is a member of the CYP84 family with important biological functions. This gene catalyzes the conversion of ferulic acid, coniferyl alcohol, and coniferaldehyde into 5-hydroxyferulic acid, 5-hydroxyconiferyl alcohol, and 5-hydroxyconiferaldehyde, respectively, making it a key enzyme regulating S-type lignin synthesis (Humphreys and Chapple, 2002). To date, the F5H gene has been cloned from various plants including *Arabidopsis thaliana* (Franke et al., 2000), *Angelica sinensis* (Wen, 2015), *Populus tomentosa* (Chen et al., 2015), *Linum usitatissimum* (Wang, 2009), and *Brassica campestris* (Li et al., 2013), but no reports have documented F5H cloning in tartary buckwheat.

Wu (2020) investigated husk composition in thick- and thin-husk tartary buckwheat, finding that lignin content in thick-husk varieties was higher than in thin-husk varieties. Our research group previously performed transcriptome sequencing and comparative analysis of thick- and thin-husk samples from an F2 population derived from a cross between 'Yunqiao 1' and 'rice buckwheat' (unpublished data), revealing that most lignin biosynthesis pathway genes showed higher expression in thick-husk than thin-husk tartary buckwheat. This suggests that thick husk phenotype may result from accumulated lignin. Transcriptome sequencing indicated that F5H gene expression was significantly higher in thick-husk compared to thin-husk varieties. To investigate the role of F5H in lignin synthesis in tartary buckwheat, this study employed RT-PCR to clone

the FtF5H gene from both 'Yunqiao 1' and 'rice buckwheat', conducted bioinformatics analysis and real-time quantitative PCR validation to provide a foundation for further functional studies.

## Materials and Methods

### 1.1 Plant Materials

Thick-husk tartary buckwheat material 'Yunqiao 1'is a cultivar developed by the Biotechnology and Genetic Germplasm Resources Research Institute of Yunnan Academy of Agricultural Sciences. Thin-husk material 'rice buckwheat' is a local Yunnan variety. Husks were stripped from different developmental stages of both varieties and pooled for gene cloning. For quantitative PCR analysis, tissues were collected during the fruiting stage, including leaves, flowers, stems, and husks from both 'Yunqiao 1' and 'rice buckwheat'.

### 1.2 RNA Extraction and First-Strand cDNA Synthesis

Total RNA was extracted using the Trizol reagent kit according to the manufacturer's instructions. RNA integrity was assessed by 1.5% agarose gel electrophoresis. For first-strand cDNA synthesis, 5 L total RNA, 1 L random primers, and 1 L $ddH_2O$ were combined in a 0.2 mL PCR tube, incubated at 70°C for 5 min, then placed on ice for 2 min. After centrifugation, 2.0 L 5× First-Strand Buffer, 0.5 L 10 mmol·$L^{-1}$ dNTPs, 0.25 L RNase inhibitor, and 0.25 L Reverse Transcriptase were added to a final volume of 10.0 L, followed by incubation at 42°C for 60 min and 72°C for 10 min.

### 1.3 FtF5H Gene Cloning

Four specific primers were designed based on the nucleotide sequence of ferulate 5-hydroxylase (FtF5H) obtained from transcriptome sequencing (Table 1). The RT-PCR reaction mixture contained 1 L cDNA template, 0.2 L dNTPs (10 mmol·$L^{-1}$), 12.5 L 2×GC Buffer I, 0.2 L Taq DNA polymerase (5 U· $L^{-1}$), 10.1 L $ddH_2O$, and 0.5 L each of forward and reverse primers (10 µmol·$L^{-1}$). PCR conditions were: 95°C pre-denaturation for 3 min; 33 cycles of 94°C denaturation for 30 s, 58°C annealing for 30 s, and 72°C extension for 90 s; followed by final extension at 72°C for 7 min. PCR products were analyzed by 1% agarose gel electrophoresis, target bands were excised and purified, and sent to Sangon Biotech (Shanghai) for sequencing.

### 1.4 Bioinformatics Analysis of Cloned Genes

The NCBI ORF finder and Conserved Domains database were used to analyze the open reading frame and conserved functional domains of the sequenced cDNA. ProtParam was employed for physicochemical property analysis, ProtScale for hydrophilicity/hydrophobicity analysis, TMHMM Server v. 2.0 for transmembrane domain prediction, SignalP 5.0 Server for signal peptide

prediction, NetPhos 3.1 Server for phosphorylation site prediction, and Psort for subcellular localization. Secondary and tertiary structures were predicted using SOPMA and SWISS-MODEL. Multiple sequence alignment of encoded proteins was performed using DNAMAN software, and neighbor-joining phylogenetic trees were constructed with MEGA 6.0 (Chen et al., 2017).

### 1.5 Real-Time Quantitative PCR (qRT-PCR) Analysis

Total RNA was extracted from eight tissue samples (leaves, flowers, stems, and husks from both 'Yunqiao 1' and 'rice buckwheat' ) using the Trizol reagent kit and reverse-transcribed into first-strand cDNA for qRT-PCR analysis, with three technical replicates per sample. The qRT-PCR reaction mixture contained 10 L 2×SG Fast qPCR Master Mix, 2 L cDNA template, 0.4 L each of forward and reverse primers, and 7.2 L dd$H_2O$. The amplification program was: 95°C for 3 min, followed by 45 cycles of 95°C for 5 s and 60°C for 30 s. Relative expression levels of FtF5H were calculated using the 2^(-ΔΔCt) method with tartary buckwheat H3 (JF769134.1) as the reference gene. Primers used are listed in Table 1.

## Results

### 2.1 Tartary Buckwheat Total RNA Extraction and FtF5H Gene Cloning

Total RNA extraction yielded clear bands on 1.5% agarose gel with brightness ratios between 1:1 and 2:1, indicating minimal degradation and suitability for subsequent cloning. RT-PCR successfully cloned the F5H gene from both 'Yunqiao 1' and 'rice buckwheat' , with products analyzed by 1.0% agarose gel electrophoresis (Figure 1 [Figure 1: see original paper]). Sequencing and assembly revealed identical nucleotide sequences for the F5H gene from both varieties, which matched the transcriptome sequencing data. ORF finder analysis confirmed a complete open reading frame of 1,395 bp encoding 464 amino acids in both sequences. This gene was designated FtF5H (GenBank accession number: MW455111). The nucleotide sequence translated into amino acid sequence is shown in Figure 2 [Figure 2: see original paper].

#### 2.2.1 Physicochemical Property Analysis

ProtParam analysis predicted the FtF5H protein molecular formula as $C_{2346}H_{3677}N_{637}O_{680}S_{28}$, with a theoretical molecular weight of 52,583.54 Da, grand average of hydropathicity (GRAVY) of -0.234, and instability index (II) of 36.57, indicating a stable protein. The aliphatic index was 82.65, with 63 negatively charged residues (Asp+Glu) and 56 positively charged residues (Arg+Lys). The theoretical isoelectric point was 5.76, suggesting an acidic protein. These characteristics indicate FtF5H is a hydrophilic, stable, acidic protein.

Conserved domain analysis identified a P450 conserved domain, classifying FtF5H as belonging to the P450 superfamily (Figure 3 [Figure 3: see original paper]). ProtScale analysis using the Hphob./Kyte & Doolittle scale revealed the highest score of 1.967 at amino acid position 256 (most hydrophobic) and the lowest score of -2.856 at position 362 (most hydrophilic). NetPhos 3.1 Server prediction identified 36 amino acid sites with scores above the threshold of 0.5, including 22 serine (Ser), 13 threonine (Thr), and 1 tyrosine (Tyr) residues, which may undergo phosphorylation to regulate protein activity and function.

SignalP 5.0 Server analysis of F5H proteins from ten plant species (*Fagopyrum tataricum*, *Malus domestica*, *Pyrus bretschneideri*, *Camptotheca acuminata*, *Oryza sativa*, *Prunus persica*, *Panicum virgatum*, *Prunus avium*, *Chenopodium quinoa*, and *Zostera marina*) predicted no signal peptides, indicating these are non-secretory proteins (Table 2 ). This suggests F5H genes are synthesized on free ribosomes and function directly in the cytoplasm without protein trafficking (Zhang et al., 2013). Psort subcellular localization prediction localized most F5H proteins to chloroplasts, except for *Panicum virgatum* which was predicted to be in the cytoskeleton. SPSS cluster analysis based on signal peptides and subcellular localization grouped dicot terrestrial plants together, while aquatic monocots (*Oryza sativa*, *Zostera marina*) formed a separate cluster, with *Panicum virgatum* alone, suggesting functional mechanism differences.

TMHMM prediction revealed that tartary buckwheat, *Panicum virgatum*, and *Zostera marina* lack transmembrane domains, while the other seven species each contain one transmembrane domain, indicating species-specific differences in F5H protein transmembrane regions.

### 2.2.2 FtF5H Structure Prediction

Secondary structure prediction by NPSA indicated FtF5H is rich in structural elements: random coil (36.21%), $\alpha$-helix (46.12%), extended strand (12.28%), and $\beta$-turn (5.39%) (Figure 5 [Figure 5: see original paper]). Tertiary structure modeling using SWISS-MODEL (Figure 6 [Figure 6: see original paper]A) showed 33.49% sequence similarity with template 5ylw.1.A for amino acids 5-455 of FtF5H, with a QMEAN score of -2.83 and GMQE (Global Model Quality Estimation) value of 0.71. The GMQE value ranges from 0-1, with values closer to 1 indicating better model quality, suggesting high reliability. Comparative analysis of protein three-dimensional structures from multiple species revealed typical functional domains: an iron protoporphyrin domain (Figure 6B) and a manganese ion domain (Figure 6C) (Zhang et al., 2015).

### 2.2.3 Amino Acid Structure Analysis and Phylogenetic Tree Construction

NCBI BlastP search identified homologous amino acid sequences showing 73.12%-77.42% identity with FtF5H. Multiple sequence alignment using DNA-

MAN revealed 84.79% similarity between tartary buckwheat FtF5H and F5H proteins from other plants, including the heme-binding domain FxxGxxxCxG (Figure 7 [Figure 7: see original paper]), where the conserved cysteine serves as the fifth ligand for heme iron (Chapple, 1998), confirming FtF5H belongs to the P450 family.

A neighbor-joining phylogenetic tree constructed with MEGA 6.0 showed tartary buckwheat FtF5H clustered most closely with *Chenopodium quinoa* CYP84A1. Previous studies reported that *Arabidopsis* ferulate-5-hydroxylase AtCYP84A1 expression correlates with quantitative and developmental regulation of lignin biosynthesis (Ruegger et al., 1999). In this study, tartary buckwheat FtF5H clustered with CYP84A1 from *Arabidopsis* and other plants (Figure 8 [Figure 8: see original paper]), suggesting its involvement in lignin synthesis and regulation in tartary buckwheat (Zhang et al., 2019).

### 2.2.4 Relative Quantitative Analysis of FtF5H Gene

qRT-PCR analysis of FtF5H expression levels in different organs of thick- and thin-husk tartary buckwheat revealed tissue-specific expression patterns. In thick-husk 'Yunqiao 1' , relative expression increased progressively from leaves to flowers to stems to husks. In thin-husk 'rice buckwheat' , the order was stems < husks < flowers < leaves. Comparisons of the same organ showed that, except for leaves, FtF5H expression in 'Yunqiao 1' was significantly higher than in 'rice buckwheat' across all tissues, with extremely significant differences (Figure 9 [Figure 9: see original paper]). Notably, FtF5H expression in thick husks was five-fold higher than in thin husks, consistent with previous transcriptome results.

## Discussion and Conclusion

Tartary buckwheat is a rare medicinal and edible crop with extremely high medicinal value and nutritional benefits. A major research focus is transferring the thin-husk trait into cultivated varieties to breed new tartary buckwheat varieties that are both easy to dehull and high-yielding with excellent quality. Song et al. (2019) investigated the relationship between cellulose and lignin content variations and dehulling efficiency, finding that decreased dehulling efficiency (thick husk) correlated with reduced lignin and increased cellulose. In contrast, Wu (2020) found that thin-husk tartary buckwheat had significantly lower cellulose and lignin content than thick-husk varieties. Our transcriptome sequencing revealed that most lignin biosynthesis pathway genes showed lower expression in thin-husk compared to thick-husk tartary buckwheat. To further explore the molecular mechanism of thin-husk formation, we cloned the F5H gene from both husk types, finding identical sequences that matched transcriptome data.

F5H is a key enzyme gene regulating plant lignin synthesis and plays an important role in lignin formation. Bioinformatics analysis revealed that the cloned tartary buckwheat F5H protein contains a P450 conserved domain, belonging

to the P450 superfamily, consistent with findings in pomegranate (Feng et al., 2018), *Eucalyptus urophylla* (Xiao et al., 2018), and *Populus tomentosa* (Wang et al., 2014). Tartary buckwheat F5H protein possesses varying numbers of serine, threonine, and tyrosine phosphorylation sites, consistent with pomegranate studies (Feng et al., 2018). Phosphorylation can alter protein activity, thereby regulating tartary buckwheat growth and development. Multiple sequence alignment showed 84.79% similarity between tartary buckwheat FtF5H and other plant F5H amino acid sequences, indicating high conservation. Phylogenetic analysis revealed the closest relationship with *Chenopodium quinoa*, with closer affinity to other dicot F5H genes than to monocot genes from rice, *Panicum virgatum*, millet, and *Daemonorops jenkinsiana*, suggesting greater homology differences between dicot and monocot classes.

Shafrin et al. (2015) downregulated F5H in jute, finding approximately 25% reduction in acid-insoluble lignin content throughout the stem and 12%-15% reduction in fiber lignin compared to non-transgenic plants. Xiao et al. (2018) reported that EuF5H showed highest expression in semi-lignified stems and lowest in young stems of *Eucalyptus urophylla*. In oilseed rape, lignin content in roots, rhizomes, and stems positively correlated with lodging resistance, with F5H expression significantly higher in lodging-resistant varieties at the bolting and flowering stages (Li and Qi, 2011). In this study, FtF5H expression in stems, flowers, and husks of thick-husk 'Yunqiao 1' was higher than in thin-husk 'rice buckwheat' , with extremely significant differences among organs. The five-fold higher expression in thick husks suggests that low FtF5H expression in thin-husk tartary buckwheat may be associated with reduced lignin synthesis and accumulation.

Takeda et al. (2017) found that OsF5H1 expression is a major factor controlling S/G lignin composition in rice cell walls. Xu et al. (2015) demonstrated that PbF5H participates in regulating S-lignin monomer synthesis and affects the G/S ratio in pear fruit. Tetreault et al. (2020) showed that overexpression of sorghum F5H (SbF5H) increased S-lignin content. Lignin synthesis involves a complex metabolic network regulated by multiple genes. As F5H is a multigene family, not all members function in G/S conversion. Therefore, the specific function of FtF5H obtained in this study within the lignin biosynthesis pathway

*Note: Figure translations are in progress. See original paper for figures.*