Initial Characterization and Expression Pattern Analysis of Tobacco (*Nicotiana Tabacum*) HMGS Gene

Meiwei Zhao[1], Tao Zhang[3], Lei Yang[3], Hongtao Feng[3], and Zhengxiong Zhao[2*]

[1]College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, China
[2]College of Resources and Environment, Yunnan Agricultural University, Kunming 650201, China
[3]Research and Development Center, China Tobacco Yunnan Industrial Co.,Ltd, Kunming 650106, China

Abstract. 3-hydroxy-3-methylglutaryl coenzyme A synthase (HMGS) is a member of condensing enzymes that catalyze a Claisen-like condensation reaction.The tobacco (nicotiana tabacum) HMGS gene was firstly characterized using the rapid amplification of cDNA ends methods based on one tobacco EST. The full-length tobacco HMGS gene mRNA was 1,773bp containing a 1389 bp open reading frame, which encodes a protein of 462 amino acids. Sequence analysis revealed that the HMGS of tobacco shares high homology with the HMGS of nicotiana tomentosiformis (96%), nicotiana attenuata (95%), Nicotiana sylvestris (95%), nicotiana benthamiana(94%), solanum lycopersicum(94%), solanum tuberosum(93%) and withania somnifera(93%). Results also showed that tobacco HMGS gene has a closer genetic relationship with the HMGS gene of withania somnifera. Tissue expression profile analysis revealed that the tobacco HMGS gene was highly expressed in flower, but moderately expressed in leaf and stem, and weakly expressed in root. Our experiment established the foundation for further research on this tobacco gene.

1 Introduction

3-hydroxy-3-methylglutaryl coenzyme A synthase (HMGS) is a member of condensing enzymes that catalyze a (decarboxylating or non-decarboxylating) Claisen-like condensation reaction. Members of these condensing enzymes are share strong structural similarity, and are involved in the synthesis and degradation of fatty acids, and the production of polyketides, a diverse group of natural products[1]. In Arabidopsis, HMGS is also known as FKP1 which is an enzyme of the mevalonate (MVA) pathway involved in biosynthesis of isoprenoids such as sterols. Knockdown of FKP1 showed that the mutation affected the development of tapetum-specific lipid-containing organelles and suggested that FKP1 is required for organelle development in tapetal cells and pollen coat formation[2].

Although HMGS play important roles in involved in the synthesis and degradation of fatty acids and polyketides, until today, the tobacco full-length HMGS gene has not been reported yet. In present experiment, we will isolate the complete mRNA sequences of this tobacco gene, subsequently perform some necessary sequence analysis and tissue expression analysis for this gene. These will establish the primary foundation of understanding this tobacco gene.

2 Material and mehods

2.1 Samples collection, RNA extraction and first-strand cDNA synthesis

Tobacco plants (Chinese commercial variety Yunyan 85) were grown in a naturally lit glasshouse with normal irrigation and fertilization. The tissues including leave, stem, root, flower were harvested and immediately frozen in liquid nitrogen and stored at -80°C[3]. Total RNA extraction and first-strand cDNA synthesis for these tissue samples were performed as the methods describe by Li et al.[4].

2.2 5'and 3'-RACE

5'- and 3'-RACE were performed as the instructions of BD SMARTTM RACE cDNA Amplification Kit (BD science, USA). For the tobacco HMGS gene, the gene specific primers (GSPs) were designed based on the coding sequence information from potato HMGS gene and its highly homologous tobacco EST sequence: FG166555.

5'-RACE GSP:
5'-ACAAAATAGGCTGCATCAGCAATCG-3'
3'-RACE GSP:
5'-TACAAGCCCATCCTCGACAGTGAAT-3'.

RACE touchdown PCR were carried out with 5 cycles of 94°C: 30 sec and 72°C: 3 min, followed by 5 cycles of

* Corresponding author: zhaozx0801@163.com

94°C: 30 sec, 66°C: 30 sec and 72°C: 3 min, finally with 25 cycles of 94°C: 30 sec, 67°C: 30 sec and 72°C: 3 min to terminate reaction. These RACE PCR products were then cloned into PMD18-T vector (TaKaRa, China) and sequenced bidirectionally with the commercial fluorometric method. At least five independent clones were sequenced for each PCR product.

2.3 Quantitative real time PCR (qRT-PCR) for tissue expression profile analysis

qRT-PCR for evaluating the level of mRNA for HMGSgene was performed on the ABI Prism 7300 Sequence Detection Systems (Applied Biosystems, Foster City, CA, USA). PCR reactions for each sample were carried out in 25µl reaction volume containing 1µl SYBR Green real-time PCR Master Mix, 100 ng cDNA template and 200 nM each primer. Conditions for real-time PCR were: an initial denaturation at 95 °C for 3 min, 40 cycles of 95 °C for 15 s, 52°C for 15 s (Table 1) and 72°C for 20 s. For each sample, reactions were set up in triplicate to ensure the reproducibility of the results. The gene relative expression levels were quantified relative to the expression of the reference gene, actin (GenBank Accession No. GQ339768) by employing the $2^{-\Delta\Delta Ct}$ value model [3,5].

2.4 Sequence analysis

The cDNA sequence prediction was conducted using GenScan software (http://genes.mit.edu/GENSCAN.html).

The protein prediction and analysis were performed using the Conserved Domain Architecture Retrieval Tool of BLAST at the National Center for Biotechnology Information (NCBI) server (http://www.ncbi.nlm.nih.gov/BLAST) and the Clustal Omega software (http://www.ebi.ac.uk/Tools/msa/clustalo/).

The theoretical isoelectric point (pI) and molecular weight (Mw) of the deduced protein of the tobacco gene was computed using the Compute pI/Mw Tool (http://www.expasy.org/tools/pi_tool.html).

3 Results

3.1 RACE results for tobacco HMGS gene

For tobacco HMGS gene, through 5'-RACE, one PCR product of 834bp was obtained. The 3'-RACE product was 1,097bp. These products were then cloned to T-vector and sequenced. Taken together, a 1,773-bp cDNA complete sequence was finally obtained (Figure 1).

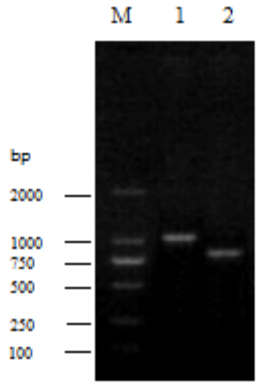

Figure 1. RACE results for tobacco HMGS gene. M DL2000 DNA markers; 1,3'-RACE product for tobacco HMGS gene; 2, 5'-RACE product for tobacco HMGS gene.

3.2 Sequence analysis

These cDNA nucleotide sequence analysis using the BLAST software at NCBI server (http://www.ncbi.nlm.nih.gov/BLAST) revealed that this gene was not homologous to any of the known tobacco gene and it was then deposited into the Genbank database (Accession number: KJ001151).

The sequence prediction was carried out using the GenScan software and results showed that the 1,773-bp cDNA sequence represents one single gene which encodes 462 amino acids (Figure 2). The pI of tobacco HMGS is 5.87. The molecular weight of this putative protein is 50978.09.

Further BLAST analysis of this protein revealed that tobacco HMGS has high homology with the HMGS of nicotiana tomentosiformis (Accession number: XP_009614165, 96%), nicotiana attenuata (Accession number: XP_019225666, 95%), Nicotiana sylvestris (Accession number: XP_009776168, 95%), nicotiana benthamiana(Accesession number: BAR94038, 94%), solanum lycopersicum(Accesession number: NP_001234846, 94%), solanum tuberosum(Accesession number: XP_006358385, 93%) and withania somnifera(Accesession number: AOX15270, 93%) (Figure 3). Its conserved domain was identified as HMG_CoA_synt_C superfamily .

The 3-D structural evidence of the putative conserved domain is also presented in figure 4. Based on the results of the alignment of different species of HMGS proteins, a phylogenetic tree was constructed using the Clustal Omega software, as shown in Figure5. The phylogenetic tree analysis revealed that the tobacco HMGS gene has a closer genetic relationship with that of withania somnifera.

Table1. qRT-PCR primers for tobacco HMGS, actin genes and annealing temperature

Gene	Primer sequence	Ta/ °C	Length/(bp)
HMGS	Forward : 5'- TGGGATGGACGCTATGGA--3' Reverse: 5'-CTGTCGAGGATGGGCTTG-3'	55	191
Actin	Forward :5'- CCATTCTTCGTTTGGACCTT -3' Reverse: 5'- TTCTGGGCAACGGAACCT-3'	56	257

```
AGACTTAGAGAGAGAAACAGTGTGTGTGTGTGTGTTTTTTTCGATCTGTCCATTTTACGTTTAGTAATACACAGAAGAAATATCAGAAAATGCAGCTCAACCGAAG
                                                                                    M  A  A  Q  P  K
AATGTCGGAATTCTCGCCGTCGAAATTTACTTCCCTCCTACTTGCCTCCCACAGGAAGCATTGGAGGCTCATGATGGAGCAAGCAAAGGAAAATACACAATTGGTCTT
 N  V  G  I  L  A  V  E  I  Y  F  P  P  T  C  L  P  Q  E  A  L  E  A  H  D  G  A  S  K  G  K  Y  T  I  G  L
GGACAAGATTGTATGGGCTTTTGCACTGAGGTTGAAGATGTTATATCAATGAGTTTGACAGCAGTTACTTCCCTTCTGGAGAAGTATGAGATTGATCCTAAGCAAATT
 G  Q  D  C  M  G  F  C  T  E  V  E  D  V  I  S  M  S  L  T  A  V  T  S  L  L  E  K  Y  E  I  D  P  K  Q  I
GGTCGTCTCGAGGTCGGAAGTGAGACAGTCATTGATAAGAGCAAATCCATCAAAACATTCCTCATGCAAATATTTGAGAAATGTGGAAACACTGACATTGAAGGAGTC
 G  R  L  E  V  G  S  E  T  V  I  D  K  S  K  S  I  K  T  F  L  M  Q  I  F  E  K  C  G  N  T  D  I  E  G  V
GACTCAACTAATGCATGTTATGGCGGAACTGCTGCATTGTTCAACTGCGTGAATTGGGTCGAGAGCTCTTCATGGGATGGACGCTATGGACTTGTTGTATGCACTGAC
 D  S  T  N  A  C  Y  G  G  T  A  A  L  F  N  C  V  N  W  V  E  S  S  S  W  D  G  R  Y  G  L  V  V  C  T  D
AGTGCGGTCTATGCTGAGGGAGCTGCTCGGCCAACTGGAGGAGCTGCAGCAATTGCTATGCTAGTAGGGCCGGATGCTCCTATTGTATTCGAAAGCAAGATTAGAGCT
 S  A  V  Y  A  E  G  A  A  R  P  T  G  G  A  A  A  I  A  M  L  V  G  P  D  A  P  I  V  F  E  S  K  I  R  A
AGCCATATGGCCCATGTCTATGATTTTTACAAGCCCATCCTCGACAGTGAATATCCAGTGGTTGATGGAAAGCTTTCACAAACTTGTTATCTTATGGCACTTGATTCT
 S  H  M  A  H  V  Y  D  F  Y  K  P  I  L  D  S  E  Y  P  V  V  D  G  K  L  S  Q  T  C  Y  L  M  A  L  D  S
TGCTACAAGAGCTTATGCGATAAATACGAAAAACTGGAAGGCAAGCAGTTTTCGATTGCTGATGCAGCCTATTTTGTTTTCCATTCACCATACAACAAGCTTGTACAG
 C  Y  K  S  L  C  D  K  Y  E  K  L  E  G  K  Q  F  S  I  A  D  A  A  Y  F  V  F  H  S  P  Y  N  K  L  V  Q
AAGAGCACTGCTCGATTGATGTTCAATGACTTTATAAGGAATGCTAGTTCCATTGACGAGTCTACTAAAGAAAAGCTTGCACCATTTTCATCCTTAACTGGTGATGAA
 K  S  T  A  R  L  M  F  N  D  F  I  R  N  A  S  S  I  D  E  S  T  K  E  K  L  A  P  F  S  S  L  T  G  D  E
AGTTATCAAAGCCGTGATCTTGAGAAAGTATCCTGGCAAGTGGCAAAACCATTTTACGATGAGAAGGTGAAACCAGCCACATTAATACCAAAACAAGTTGGCAACATG
 S  Y  Q  S  R  D  L  E  K  V  S  W  Q  V  A  K  P  F  Y  D  E  K  V  K  P  A  T  L  I  P  K  Q  V  G  N  M
TACACTGCTTCTCTATATGCTGCTTTTGCATCACTCCTTCACAATAAGCACAACACGTTGGCTGGACAGCGGTAATGTTGTTCTCCTATGGTAGTGGATCAACTGCA
 Y  T  A  S  L  Y  A  A  F  A  S  L  L  H  N  K  H  N  T  L  A  G  Q  R  V  M  L  F  S  Y  G  S  G  S  T  A
ACAATGTTCTCACTCAAGCTTAATGAAGGTCAACATCCTTTTAGCTTGTCAAACATTGCAAGCGCGATGAATGTTGCAGAGAAGTTGAAATCAAGACACGAGTTGGTT
 T  M  F  S  L  K  L  N  E  G  Q  H  P  F  S  L  S  N  I  A  S  A  M  N  V  A  E  K  L  K  S  R  H  E  L  V
CCTGAAAAATTCGTCGAAATAATGCAACTAATGGAGCATAGATATGGGGCCAAGGATTTCGTCACAAGCAAAGATTGTAGTCTTCTAGCTCCAGGAACTTACTACCTC
 P  E  K  F  V  E  I  M  Q  L  M  E  H  R  Y  G  A  K  D  F  V  T  S  K  D  C  S  L  L  A  P  G  T  Y  Y  L
ACAGAGGTTGATTCCAAGTACAGAAGATTCTATGCCAAGAAAGCTTCTGAGAATGGACTGGTGAATGGTCACTGAAGATAGTCCCAGCACCTAAGTGTAACGCCACAAAT
 T  E  V  D  S  K  Y  R  R  F  Y  A  K  K  A  S  E  N  G  L  V  N  G  H  *
TTGCTTGTGCTCTTTTAGTTGTCAGTATAGGGAATAATGTCCTAGTAAAGTTGTTTCTATTTATATAGTGTCTTTTGTTGATTAGCTTCAATGCTTTTCTTTTTTGTATT
TCATTTTTTTTCCAATTAAAGTTTACTAATTCGATGTTTATTGTGTTTGTCTTAGAGATTGTAGATGATTGACAAATCATCAATGTAATCTTGATTGAAATTCAATGAGT
AGATTTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Figure 2. The complete mRNA of tobacco HMGS gene and its encoding amino acids. *indicates the stop codon

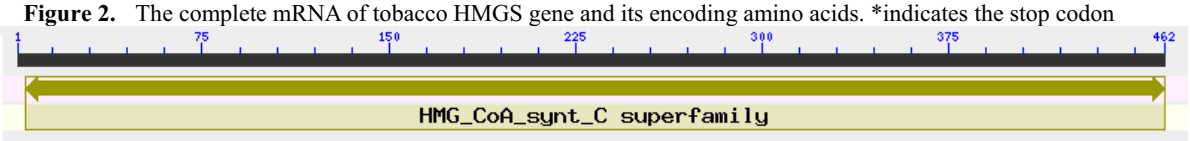

Figure 3. The putative HMG_CoA_synt_C superfamily domain of the protein encoded by tobacco HMGS gene

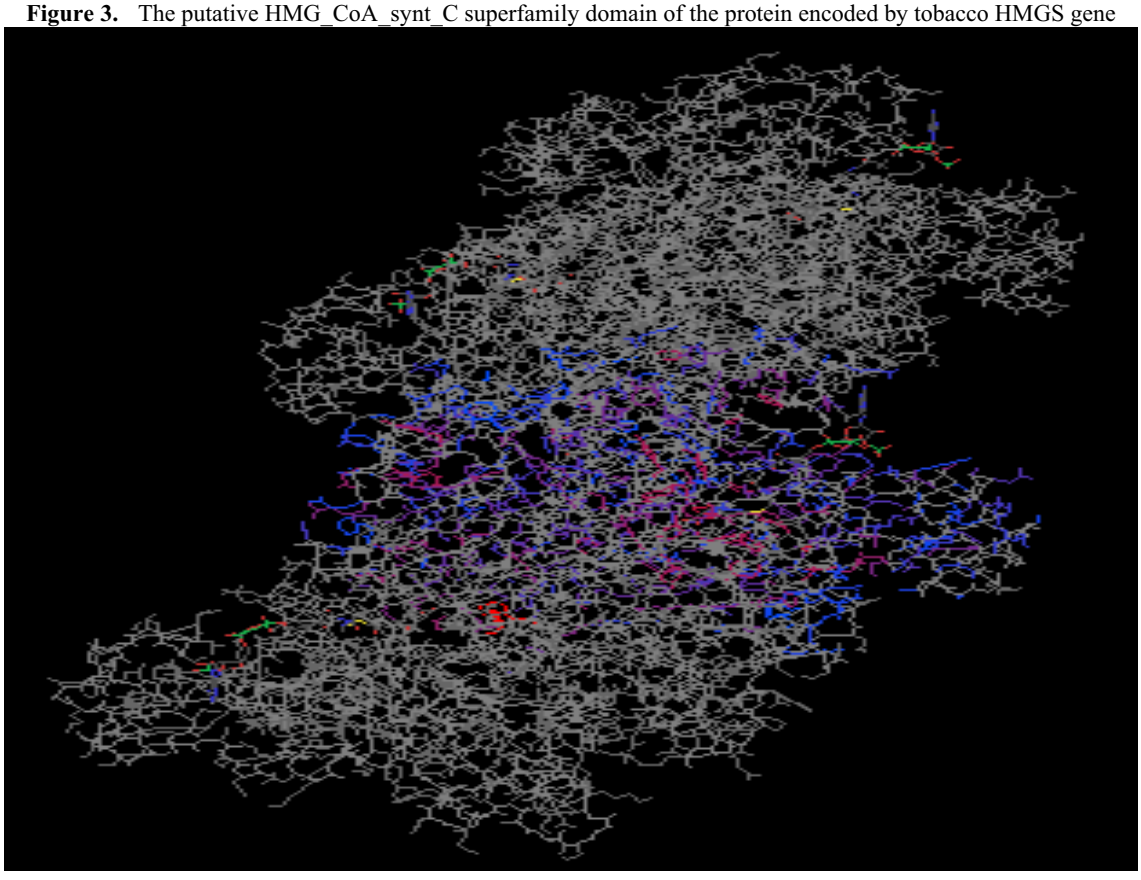

Figure 4. The 3-D structural evidence of the putative conserved domain of tobacco HMGS protein

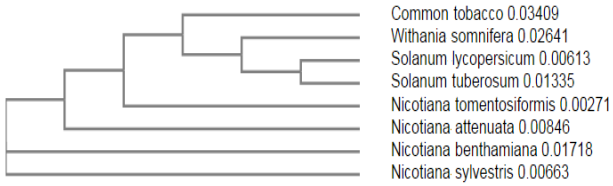

Figure 5. The phylogenetic tree for eight kinds of HMGS genes

3.3 Tissue expression profile

Tissue expression profile analysis was carried out and results revealed that the tobacco HMGS gene was highly expressed in flower, but moderately expressed in leaf and stem, and weakly expressed in root (Figure 6).

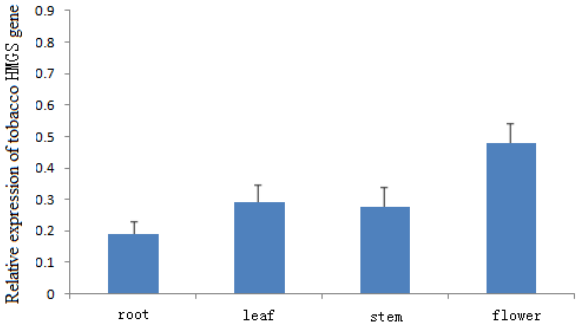

Figure 6. Expression analysis of HMGS gene mRNA in various tissues

4 Conclusions

Comparative genomics research has revealed that the extensive conservation exists in protein-coding genes of eukaryotes[6]. The HMGS gene has been reported to be conserved in human, chimpanzee, dog, cow, mouse, rat, chicken, zebrafish, fruit fly, mosquito and other animals [7,8]. From the sequence analysis of HMGS genes, it can be seen that the coding sequences of HMGS genes were also highly conserved in tobacco and some other plants. HMGS is a member of condensing enzymes which share strong structural similarity[1]. This implied that our results coincide with conclusions of Hardison[6]and Marchler-Bauer et al.[1] and are reliable.

The phylogenetic tree analysis revealed that the tobacco HMGS gene has a closer genetic relationship with that of withania somnifera. This implied that we can use withania somnifera as model organism to study the tobacco HMGS gene or use tobacco as model organism to study the withania somnifera HMGS gene.

From the tissue distribution analysis in our experiment it can be seen that this HMGS gene was obviously deferentially expressed in some tissues and was highly expressed in flower. The expression patterns of genes can often partly reveal their likely physiological functions[9]. This implied that HMGS maybe play an important role at the flowering stage of tobacco and this result also coincide with conclusion that HMGS is required for organelle development in tapetal cells and pollen coat formation[2].

In conclusion, we first cloned the full-length tobacco HMGS gene and performed necessary sequence analysis and tissue expression profile analysis. This established the primary foundation for further research on this tobacco gene.

Acknowlegement

This work was supported by The Research and Development Foundation of Yunnan Agricultural University (KX900187000) and The Basic Research Foundation of Yunnan Tobacco Industry Co. Ltd."Construction and application of R&D information platform based on market demand orientation"(2019xx01).

References

1. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Nucleic Acids Res. **45**, 3 (2017)
2. Ishiguro S, Nishimori Y, Yamada M, Saito H, Suzuki T, Nakagawa T, Miyake H, Okada K, Nakamura K, Plant Cell Physiol. **51**, 16 (2010)
3. Nian F, Zhang Y, Su X, Zou J and Zhao L, IJAST. **1**,5 (2013)
4. Li Y, Meng F, Yin J, Liu H, Si Z, Ni Z, Sun Q, Ren J, Niu H, Biochim Biophys Acta. **1779**, 9 (2008)
5. Livak KJ, Schmittgen TD, Methods. **25**, 7 (2001)
6. Hardison RC. PLoS Biol. **1** (2003)
7. Zhao L, Fan J, Xia S, Pan Y, Liu S, Qian G, Qian Z, Kang HB, Arbiser JL, Pollack BP, Kudchadkar RR, Lawson DH, Rossi M, Abdel-Wahab O, Merghoub T, Khoury HJ, Khuri FR, Boise LH, Lonial S, Chen F, Chen J, Lin R, J Biol Chem. **292**, 11 (2017)
8. L.S. Einbond, Soffritti M, Esposti DD, Park T, Cruz E, Su T, Wu HA, Wang X, Zhang YJ, Ham J, Goldberg IJ, Kronenberg F, Vladimirova A, Fundam Clin Pharmacol. **23**, 11 (2009)
9. Wang Z, Wei P, Wu M, Xu Y, Li F, Luo Z, Zhang J, Chen A, Xie X, Cao P, Lin F, Yang J, Planta. **242**, 14 (2015)