E3S Web of Conferences **292**, 03070 (2021)
NETID 2021
https://doi.org/10.1051/e3sconf/202129203070

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

Meiwei Zhao[1], Lei Yang[3], Jiacan Wu[3], Haijuan Wang[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. The complete mRNA sequence of one tobacco (nicotiana tabacum) gene—guanosine monophosphate (GMP)synthase, was amplified using the rapid amplification of cDNA ends methods. The full-length tobacco GMP synthase gene mRNA was 2,127bp containing a 1,617 bp open reading frame, which encodes a protein of 538 amino acids. Sequence analysis revealed that the GMP synthase of tobacco shares high homology with the GMP synthase of wood tobacco(99%), nicotiana attenuata(99%), nicotiana tomentosiformis(99%), potato(92%), Lycopersicon pennellii(92%), lycopersicon esculentum(92%), capsicum annuum(91%), capsicum chinense(91%) and capsicum baccatum(90%). BLAST analysis within the tobacco high throughout genomic sequences database revealed that this gene has 5 introns and 6 exons. Results also showed that tobacco GMP synthase gene has a closer genetic relationship with the GMP synthase gene of wood tobacco. Tissue expression profile analysis revealed that the tobacco GMP synthase gene was highly expressed in leaf, but moderately expressed in root, flower and stem. Our experiment established the foundation for further research on this tobacco gene.

1 Introduction

GMP synthase or glutamine amidotransferase catalyzes the synthesis of GMP from xantosine monophosphate.This protein is a homodimer, but in some archaea it is a heterodimer composed of a glutamine amidotransferase subunit and a ATP pyrophosphatase subunit. In eucaryotes, bacteria, and some archaea the two catalytic units are encoded by a single gene, producing a two-domain-type GMP, with a GATase domain in the N-terminal half and a ATP-PPase domain in the C-terminal half. The C-terminal domain is specific to the GMP synthases. (http://www.ebi.ac.uk/interpro/entry/IPR022310)[1-4].

GMP synthetase is a glutamine amidotransferase from the de novo purine biosynthetic pathway[5]. The amidotransferase family of enzymes utilises the ammonia derived from the hydrolysis of glutamine for a subsequent chemical reaction catalyzed by the same enzyme. The ammonia intermediate does not dissociate into solution during the chemical transformations[6].Recent research reported that GMP synthase is essential for viability and infectivity of Trypanosoma brucei despite a redundant purine salvage pathway[1].

Although GMP synthase play important roles in GMP biosynthesis of eucaryotes, bacteria and some archaea, until today, the tobacco GMP synthase 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℃[7]. Total RNA extraction and first-strand cDNA synthesis for these tissue samples were performed as the methods describe by Li et al.[8].

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 GMP synthase gene, the gene specific primers (GSPs) were designed based on the coding sequence information from potato GMP synthase gene and its highly homologous tobacco EST sequence: EB679319.

5’-RACE GSP:

* Corresponding author: zhaozx0801@163.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

5’- CAATTCACGAACCTCATCCTTGAAT-3’

3’-RACE GSP:

5’- AAGGCGTAACAGAACCTGAAATGAA -3’.

RACE touchdown PCRs were carried out with 5 cycles of 94°C: 30 sec and 72°C: 3 min, followed by 5 cycles of 94°C: 30 sec, 64°C: 30 sec and 72°C: 3 min, finally with 25 cycles of 94°C: 30 sec, 63°C: 30 sec and 72°C: 3 min to terminate reaction. These RACE PCR products were then cloned into T-vector (TaKaRa, China) and sequenced.

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

qRT-PCR for evaluating the level of mRNA for GMP synthase gene 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 [7,9].

2.4 Sequence analysis

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

The protein prediction and analysis were performed u sing the Conserved Domain Architecture Retrieval Tool of BLAST at the National Center for Biotechnology Info rmation (NCBI) server (http://www.ncbi.nlm.nih.gov/BL AST) and the Clustal Omega software (http://www.ebi.a c.uk/Tools/msa/clustalo/).

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

3 Results

3.1 RACE results for tobacco GMP synthase gene

For tobacco GMP synthase gene, through 5’-RACE, one PCR product of 1,265-bp was obtained. The 3’-RACE product was 1,152-bp. These products were then cloned to T-vector and sequenced. Taken together, a 2,127-bp cDNA complete sequence was finally obtained (Figure 1).

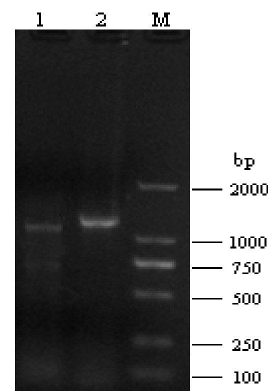

Fig 1. RACE results for tobacco GMP synthase gene. M DL2000 DNA markers; 1, 3’-RACE product for tobacco GMP synthase gene; 2, 5’-RACE product for tobacco GMP synthase gene

3.2 Sequence analysis

These cDNA nucleotide sequence analysis using the BL AST software at NCBI server (http://www.ncbi.nlm.nih. gov/BLAST) revealed that this gene was not homologou s to any of the known tobacco gene and it was then depo sited into the Genbank database (Accession number: KJ0 01142).

The sequence prediction was carried out using the Ge nScan software and results showed that the 1,278-bp cD NA sequence represents one single gene which encodes 538 amino acids (Figure 2). The pI of tobacco GMP synt hase is 6.51. The molecular weight of this putative protei n is 60109.32. BLAST analysis within the tobacco high t hroughout genomic sequences database revealed that this gene has no intron and is a single exon gene.

Further BLAST analysis of this protein revealed that tobacco GMP synthase has high homology with the GM P synthase of wood tobacco (Accession number: XP_00 9764602, 99%), nicotiana attenuata (Accession number: XP_019231785, 99%), nicotiana tomentosiformis(Acces sion number: XP_009600064, 99%), potato(Accesession number: XP_006343975 , 92%), Lycopersicon pennellii (Accesession number: XP_015085385, 92%), lycopersic on esculentum (Accesession number: XP_004245599, 9 2%), capsicum annuum (Accesession number: XP_0165 67152 , 91%), capsicum chinense (Accesession number: PHU28017, 91%) and capsicum baccatum (Accesession number: PHT57714 , 90%). Its conserved domain was id entified as GATase superfamily and GMP_synthase_C (Figure 3).

The 3-D structure of the putative conserved domain a lso showed that C-terminal of tobacco the GMP synthase contains the GMP_synthase_C domain (Figure 4).

Based on the results of the alignment of different spe cies of GMP synthase proteins, a phylogenetic tree was c onstructed using the Clustal Omega software, as shown i n Figure 5. The phylogenetic tree analysis revealed that t he tobacco GMP synthase gene has a closer genetic relati onship with that of wood tobacco.

Table 1. qRT-PCR primers for tobacco GMP synthase, actin genes and annealing temperature

Gene	Primer sequence	Ta/ °C	Length/(bp)
GMP synthase	Forward : 5'- CTAATCACCCGCCGAATC-3' Reverse: 5'-GCCACAACCTCAAACCCT -3'	55	410
Actin	Forward :5'-CCATTCTTCGTTTGGACCTT -3' Reverse: 5'- TTCTGGGCAACGGAACCT-3'	56	257

```
ATATAAACCCTAGCCTCCATTTCTCTCTATCTTTCTTCCTCCCACCCACCACCCACCCTCCCCTAGCAATGGAACCTCAAACACAGGCGAAGAAATCAAACCTCGTACTAATCCTAGACTACGGTTCTCAGTACACTCAC
                                                                     M  E  P  Q  T  Q  A  K  K  S  N  L  V  L  I  L  D  Y  G  S  Q  Y  T  H
CTAATCACCCGCCGAATCCGAAGCCTATCAATTTTCTCACTCACCATTAACGGCACCTCTTCGTTAGACTCCATAAAAGAACTCGACCCACGTGTCATTATCCTCTCGGGTGGACCCCACAGCGTCCACGCTGACGGC
 L  I  T  R  R  I  R  S  L  S  I  F  S  L  T  I  N  G  T  S  S  L  D  S  I  K  E  L  D  P  R  V  I  I  L  S  G  G  P  H  S  V  H  A  D  G
GCACCGTGTTTCCCACCTGGGTTCATCGAATACGTCGAGTCACGTGGGATTCACGTGTTGGGTATATGTTATGGGCTGCAGTTGATTGTTCAGAAACTTGGCGGGGTTGTGAAAATTGGAGAGAAACATGAGTATGGG
 A  P  C  F  P  P  G  F  I  E  Y  V  E  S  R  G  I  H  V  L  G  I  C  Y  G  L  Q  L  I  V  Q  K  L  G  G  V  V  K  I  G  E  K  H  E  Y  G
AGAATGGAAATTGAGGTTGGAAAGAATGTTGTTGGGGGGTTGTTTGGGAATACGGAAATTGGTGATAAACAGGTGGTTTGGATGAGCCACGGTGATGAGGCTGTGAAATTGCCGGAAGGGTTTGAGGTTGTGGCGAGG
 R  M  E  I  E  V  G  K  N  V  V  G  G  L  F  G  N  T  E  I  G  D  K  Q  V  V  W  M  S  H  G  D  E  A  V  K  L  P  E  G  F  E  V  V  A  R
AGTAGTCAGGGTGCTGTTGCTGCTATTGAGAATCGGGAACGGAGGTTTTATGGGCTGCAGTATCATCCCGAGGTAACGCACTCGACTGAAGGGATGAGAACATTAAGACACTTTCTGTTTGATGTATGTGGGATTACA
 S  S  Q  G  A  V  A  A  I  E  N  R  E  R  R  F  Y  G  L  Q  Y  H  P  E  V  T  H  S  T  E  G  M  R  T  L  R  H  F  L  F  D  V  C  G  I  T
GCTGGCTGGAAGATGGAAGATGTTCTGGAGGAAGAAATAAAAGTTATCAAAGGCATGGTTGGACCTGAAGATCACGTGATTTGTGCTTTATCTGGTGGTGTTGATTCCACAGTTGCAGCTAAATTGGTACACAAGGCT
 A  G  W  K  M  E  D  V  L  E  E  E  I  K  V  I  K  G  M  V  G  P  E  D  H  V  I  C  A  L  S  G  G  V  D  S  T  V  A  A  K  L  V  H  K  A
ATCGGGGACAGGCTTCACTGTGTTTTTGTCGATAATGGTCTATTAAGGTATAAGGAGAGAGAAAGGGTGATGGACCTCTTTGAGAAGCGCCTCCATTTGCCTGTTACCTGTGTCGATGCTACAGAAGAATTTCTCAGC
 I  G  D  R  L  H  C  V  F  V  D  N  G  L  L  R  Y  K  E  R  E  R  V  M  D  L  F  E  K  R  L  H  L  P  V  T  C  V  D  A  T  E  E  F  L  S
AAACTAAAAGGCGTAACAGAACCTGAAATGAAGAGGAAAATAATTGGGAAGGAGTTCATCAACATATTTGATCTTTTTGCCCATGATGTGGAGAAAAAAGTAGGGAAAAAACCTACTTACCTAGTCCAAGGAACCTTG
 K  L  K  G  V  T  E  P  E  M  K  R  K  I  I  G  K  E  F  I  N  I  F  D  L  F  A  H  D  V  E  K  K  V  G  K  K  P  T  Y  L  V  Q  G  T  L
TATCCTGATGTAATAGAGTCTTGTCCTCCGCCTGGAAGTGGAAGAACACATTCTCATACAATCAAGAGTCATCATAATGTTGGAGGTCTTCCTAAGGACATGAAGCTGAAGCTCATAGAGCCACTGAAACTTCTATTC
 Y  P  D  V  I  E  S  C  P  P  P  G  S  G  R  T  H  S  H  T  I  K  S  H  H  N  V  G  G  L  P  K  D  M  K  L  K  L  I  E  P  L  K  L  L  F
AAGGATGAGGTTCGTGAATTGGGAAAGATTTTGGATATATCTGAGGACTTTCTTAAACGCCACCCGTTCCCTGGGCCCGGGCTCGCTGTACGAATTCCAGGTGATGTCACAGCGGGGAATTCCTTGGATATTCTTCGT
 K  D  E  V  R  E  L  G  K  I  L  D  I  S  E  D  F  L  K  R  H  P  F  P  G  P  G  L  A  V  R  I  P  G  D  V  T  A  G  N  S  L  D  I  L  R
CAGGTTGATGAGATCTTCATTCAATCAATCAGAGATGCTAAAATCTATGATGAAATATGGCAAGCTTTTGCTGTCTTCTTACCAGTGAAAACTGTTGGAGTACAAGGAGACCAAAGAACACATTCCCATGCTGTTGCA
 Q  V  D  E  I  F  I  Q  S  I  R  D  A  K  I  Y  D  E  I  W  Q  A  F  A  V  F  L  P  V  K  T  V  G  V  Q  G  D  Q  R  T  H  S  H  A  V  A
CTTAGAGCAGTCACAAGTCAAGATGGAATGACTGCAGACTGGTACTACTTTGATTTCAAGTTTCTTGACGACGTATCAAGAAAGATCTGCAATAGTGTTCGTGGTGTAAATCGAGTTCTTCTAGATATTACATCAAAG
 L  R  A  V  T  S  Q  D  G  M  T  A  D  W  Y  Y  F  D  F  K  F  L  D  D  V  S  R  K  I  C  N  S  V  R  G  V  N  R  V  L  L  D  I  T  S  K
CCTCCATCAACAATCGAGTGGGAATAATTTGTTATAAAGAATGCTATATTTGGTGACCAAAGTAGGATTCTTTTGTGATTTTTGGTGCATAACAAAAAGGAAGAAAATCATGATAGAAATTTAGGTCCTTTTGTTATGTG
 P  P  S  T  I  E  W  E  ^
GTAGAACTGGTTCTTGGGTAATTATGTGCAATGCTCTCGACAATTTTGTACGCGGGGAATGCATTGGATGGATGCCCGGGCATTGAGAAGGAAGGACGCTTTCAGAGGCGAAAGGCCATGGGGAGATACCGTCTGTGATC
CATGGATCTCCGATCGGGAAACCGTATCCAAGCTCCGTGGCTAGTTTGCGCTCTTTGGACTTTTCAAACTTAGCGAACTGAAACATCTTAGTAGCTAAAGGAAGGGAAATCAACCGAGACCCCGTTAGTAGCGGCGAGCG
AGAGCGGATTGGGGGTTTGAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig 2. The complete mRNA of tobacco GMP synthase gene and its encoding amino acids *indicates the stop codon

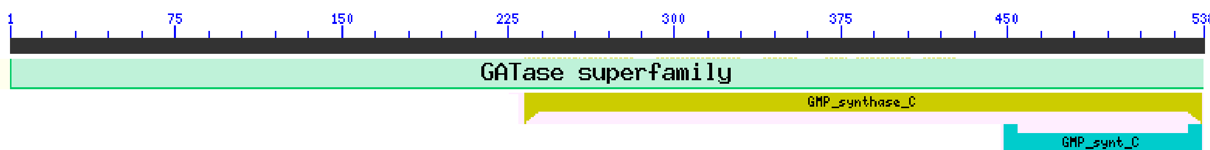

Fig 3. The putative GATase superfamily domain and GMP_synthase_C domain of the protein encoded by tobacco GMP synthase gene

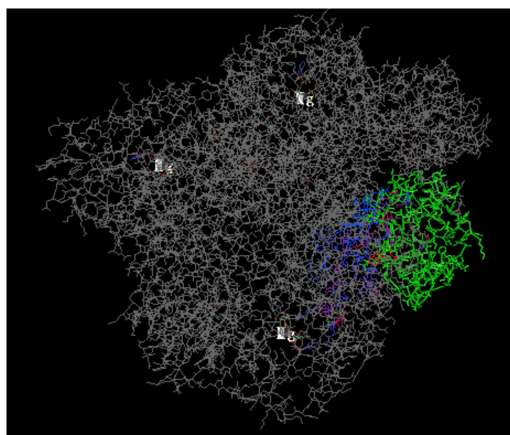

Fig 4. The 3-D structural evidence of the putative conserved GMP_synthase_C domain of tobacco GMP synthase protein

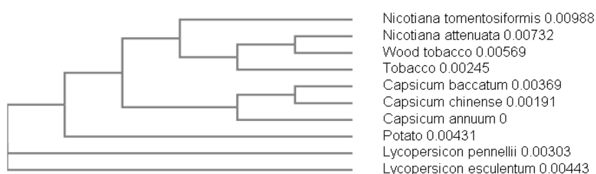

Fig 5. The phylogenetic tree for six kinds of GMP synthase gene

To obtain the genomic DNA of common tobacco GMP synthase gene, the publicly available common tobacco whole genome shotgun sequence database was screened using the coding sequences of common tobacco GMP synthase gene as seeds by BLAST analysis. A whole genome shotgun sequence (GenBank accession no. NW_015883288.1) was found to encompass the entire GMP synthase gene (8,071bp-12,578bp). The common tobacco GMP synthase gene is 4,508-bp in length and consists of 6 exons and 5 introns. All exon–intron splice junction sequences conform to the GT–AG rule (Figure 6).

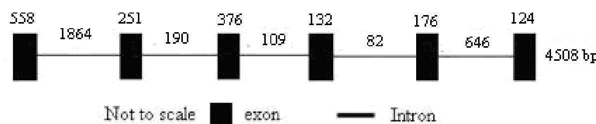

Fig 6. The genomic sequence organizations representing the coding regions of the common tobacco GMP synthase gene

3.3 Tissue expression profile

Tissue expression profile analysis was carried out and results revealed that the tobacco GMP synthase gene was highly expressed in leaf, but moderately expressed in root, stem and flower (Figure 7).

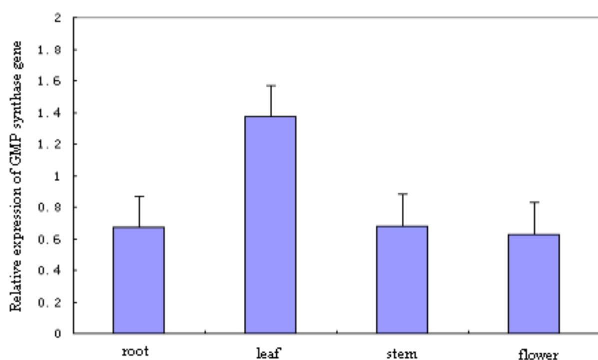

Fig 7. Expression analysis of GMP synthase gene mRNA in various tissues

4 Conclusions

Protein comparison analysis revealed that tobacco GMP synthase shares high homology with several GMP synthases. GMP synthetase is a bifunctional two-domain enzyme. a GATase domain in the N-terminal half and a ATP-PPase domain in the C-terminal half which is specific to the GMP synthases[1-4]. In our results, the putative conserved domain analysis showed that of tobacco GMP synthase contains the GATase superfamily and GMP_synthase_C conserved domains. These imply that our study results are reliable.

The phylogenetic tree analysis revealed that the tobacco GMP synthase gene has a closer genetic relationship with that of wood tobacco. This implied that we can use wood tobacco as model organism to study the tobacco GMP synthase gene or use tobacco as model organism to study the wood tobacco GMP synthase gene. From the tissue distribution analysis in our experiment it can be seen that GMP synthase gene was obviously differentially expressed in some tissues. The tobacco GMP synthase gene was highly expressed in leaf, but moderately expressed in root, stem and flower. For GMP synthase catalyzes the synthesis of GMP from xantosine monophosphate[1-4]. The suitable explanation for differential expression under current conditions is that the GMP biosynthesis was high in leaf, but moderate in root, stem and flower.

In conclusion, we first isolated the tobacco GMP synthase gene and performed necessary sequence analysis and tissue expression analysis. This established the primary foundation for further research on this tobacco gene.

Acknowlegement

This work was supported by The Research and Develop ment Foundation of Yunnan Agricultural University (KX 900187000) and The Basic Research Foundation of Yun nan Tobacco Industry Co. Ltd."Construction and applica tion of R&D information platform based on market dema nd orientation"(2019xx01).

References

1. LI Q., LEIJA C., RIJO-FERREIRA F., CHEN J., CESTARI I., STUART K., TU B.P., PHILLIPS M.A., Mol Microbiol. **97**, 15 (2015)
2. WELIN M., LEHTI Ö L., JOHANSSON A., FLODIN S., NYMAN T., TRÉSAUGUES L., HAMMARSTR ÖM.M., GRÄSLUND S., NORDLUND P., J Mol Biol. **425**, 11 (2013)
3. ZALKIN H., ARGOS P., NARAYANA S.V., TIEDEMAN A.A., SMITH J.M., J Biol Chem. **260**, 1 (1985)
4. ABBOTT J.L., NEWELL J.M., LIGHTCAP C.M., OLANICH M.E., LOUGHLIN D.T., WELLER M.A., LAM G., POLLACK S., PATTON W.A.,. Protein J. **25**, 9 (2006)
5. MARGOLIS, N., HOGAN, D., TILLY, K., ROSA, P. A., J Bacteriol. **176**, 6 (1994)
6. RAUSHEL F.M., THODEN J.B., HOLDEN H.M., Biochemistry. **38**, 9 (1999)
7. NIAN F., ZHANG Y., SU X., ZOU J., ZHAO L., IJAST. **1**, 5 (2013)
8. LI Y., MENG F., YIN J., LIU H., SI Z., NI Z., SUN Q., REN J., NIU H., Biochim Biophys Actav. **1779**, 9 (2008)
9. LIVAK K.J., SCHMITTGEN T.D., Methods. **25**, 7 (2001)