

Research Article

Open Access

Bioinformatics Analysis of Cysteine Transmembrane Gene CYS in Transgenic Tobacco

Luo Xiangyi¹, Gao Jieming², Liu Minghui¹, Cao Dong³, Zong Yuan^{2,3}, Liu Baolong³✉, Wei Le¹✉

¹ Qinghai Normal University, Xining, 810008, China

² Qinghai University, Xining, 810016, China

³ Key Laboratory of Crop Molecular Breeding of Qinghai Province, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810001, China

✉ Co-corresponding authors email: beyond_3862740@163.com; bliliu@nwipb.cas.cn

Plant Gene and Trait, 2023, Vol.14, No.4 doi: [10.5376/pgt.2023.14.0004](https://doi.org/10.5376/pgt.2023.14.0004)

Received: 07 Mar., 2023

Accepted: 14 Mar., 2023

Published: 29 Mar., 2023

Copyright © 2023 Luo et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Luo X.Y., Gao J.M., Liu M.H., Cao D., Zong Y., Liu B.L., and Wei L., 2023, Bioinformatics analysis of cysteine transmembrane gene CYS in transgenic tobacco, Plant Gene and Trait, 14(4): 1-8 (doi: [10.5376/pgt.2023.14.0004](https://doi.org/10.5376/pgt.2023.14.0004))

Abstract Overexpression of MYB transcription factor *LrAN2* in tobacco can activate up-regulation of anthocyanin biosynthetic pathways, but the expression of cysteine-rich transmembrane domain protein A (CYS) is significantly reduced. AS an important biological amino acid in plants, the biological research of CYS is rarely reported. In this study, the cloned tobacco CYS gene is 282 bp long and encoding 94 amino acids. The predicted molecular weight and isoelectric point of the encoded protein are 24 019.84 Da and 5.23 respectively, and there are no conserved regions in the sequence. The experimental results show that the protein is hydrophilic, mainly composed of α -helix (19.35%) and irregular coil (65.59%), CYS protein does not pass through the transmembrane region and acts outside the membrane. Phylogenetic tree shows that the CYS gene in tobacco was closely related to the CYS gene in pepper, followed by HTC gene and WIH2 gene in tomato. The results of this study cloned and analyzed the sequence of the coding region of CYS gene to provide a reference for the functional study of CYS gene.

Keywords Tobacco (*Nicotiana tabacum* L.); CYS; Gene cloning; Bioinformatics analysis

Tobacco (*Nicotiana tabacum* L.) is an annual plant of *Nicotiana* genus in the family of Solanaceae. As one of the important cash crops, tobacco is widely grown throughout the world. Tobacco can not only be made into economic commodities like cigarettes, but also has high medicinal value such as treating chronic and cardiovascular diseases, relieving cough and sterilization (Wang, 2016, Agriculture of Henan, (32): 57). In addition, the large amount of various amino acids in tobacco protein can produce high value-added peptide and amino acid products after process. These proteins play an important role in diet, biochemistry and medicine and has a wide range of applications (Liu et al., 2014). In the early stage of this study, a class of MYB transcription factors, *LrAN2*, which contains MYB structural domain and induces phycocyanin synthesis, was isolated from Heigouqi (*Lycium ruthenicum*). This gene clusters with MYB transcription factors related to the regulation of anthocyanin anabolism in other Solanaceae crops (Zong et al., 2019b). The expression of *LrAN2* increased with fruit development, and all tissue parts of the tobacco showed a various purple phenotype after over expression of *LrAN2* in tobacco. Transcriptome analysis of purple tobacco and green wild type showed that all structural genes involved in anthocyanin biosynthesis were activated in purple tobacco, but some genes were also down-regulated in purple tobacco, with cysteine-rich transmembrane structural domain protein A (CYS) being the most significantly down-regulated in purple tobacco (Zong et al., 2019a).

Cysteine is a conditionally essential amino acid required for the sustenance of living organisms, it is an α -amino acid and is a sugar-generating amino acid (Wang et al., 2018). Glutathione (GSH), which acts as an antioxidant in the body, is mostly composed of cysteine. Cysteine exists in most of the cellular structures of plants. Sulphur is an important nutrient in plants, and the first thing plants do when they take up and restore the Sulphur oxides is to synthesize cysteine to make it into their various metabolisms (Wang et al., 2011). Cysteine synthetase complexes can regulate cysteine synthesis better, it can be generated by O-acetyl serine thiol cleavage enzyme (OASS) with serine acetyltransferase (SAT). Cysteine protease is an important protease which plays a role in various plant

physiological responses, including seed germination, flower and leaf development, fruit ripening and legume rhizome growth, and in plant responses to various stresses (including drought, pathogen infection). It was found that the amount of mRNA in cysteine proteases increased under stresses such as salinity, low temperature and drought, as well the amount of mRNA for cysteine proteases when certain plant cells underwent programmed death. As a result, plant development will change in different degrees because of the reduction of the activity of cysteine proteases (Zhao et al., 2012).

During the ageing process of plant leaf, cysteine proteases generate large amounts of cellular nitrogen by degrading leaf proteins for recycling (Xiao et al., 2014). Cysteine proteases act centered on cysteine residues and are involved in a variety of protein hydrolysis functions in higher plants (Roberts et al., 2012). In recent years, research on cysteine proteases related to plant senescence has focused on the mRNA level and protease activity after translation (Maciel et al., 2011). Researchers have recently cloned a new senescence-associated gene *CaCp* in pepper, which belongs to the papain superfamily and is highly homologous to other known cysteine proteins. The gene plays a negative regulatory role in the defense response to salt stress and osmotic stress in pepper, thereby delaying leaf senescence. Meanwhile, two up-regulated cDNAs, SENU2 and SENU3, encoding cysteine proteases, isolated by the investigators from a tomato leaf senescence library, and SENU2 and SENU3 mRNAs were closely associated with leaf senescence (Zang et al., 2010).

The overexpression of MYB transcription factors in tobacco suppresses the expression of CYS as well. In this study, we plan to isolate the CYS genome and coding region sequences to gain a preliminary understanding of the bioinformatics basis related to the CYS gene and to prepare the preliminary groundwork for subsequent functional studies of CYS in tobacco.

1 Results and Analysis

1.1 Cloning of target gene

The experimental materials were transgenic purple tobacco (Figure 1A) and negative control tobacco Samsun (Figure 1B). Total tobacco RNA was extracted and the RNA electropherograms are as follows (Figure 2A), with clear 18S and 28S bands, and the RNA samples were ready for the next step of the experiment. We subjected the above extracted RNA to reverse transcription, and used the specific primers CYS-F: ATGAGTTACTACAATCAAC and CYS-R: TCAGAAGCATGCATCTAGGA for amplification to obtain a 282 bp size sequence of the target band, and the amplification results (Figure 2B) showed that the target band was specific.



A

B

Figure 1 Transgenic purple tobacco (A) negative control tobacco (B)

The amplified fragments were recovered and subjected to TA cloning. The results of the colony PCR assay showed that the amplified bands matched the size of the target bands, which were all around 282 bp (Figure 3).

1.2 Sequence analysis of the CYS gene

The full length of the CYS gene analyzed by ExPaSy-Protparam software was 282 bp (Figure 4). The predicted molecular weight of the protein was 24 019.84 Da, the isoelectric point is 5.23 and the instability index is 73.95, indicating that the protein was unstable. Aliphatic index is often used as a positive factor of increased protein heat

stability and is the proportion of aliphatic side such as alanine and leucine in a protein. The aliphatic index of this protein is 24.82.

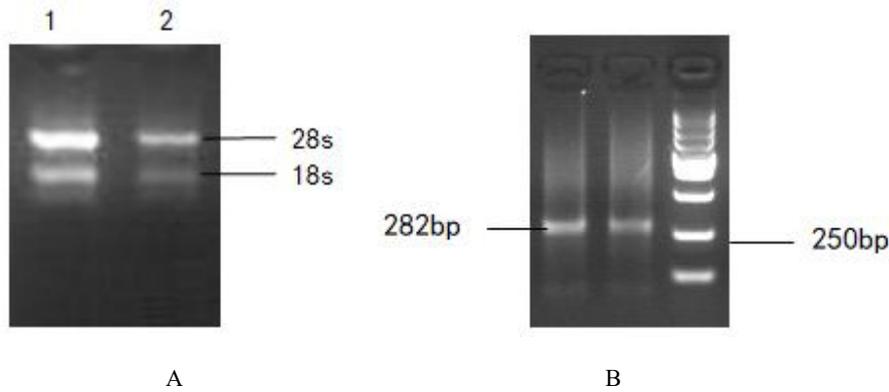


Figure 2 RNA detection map of tobacco amplification (A) amplification of tobacco cDNA (B)

Note: 1: Tobacco RNA repeat 1; 2: Tobacco RNA repeat 2

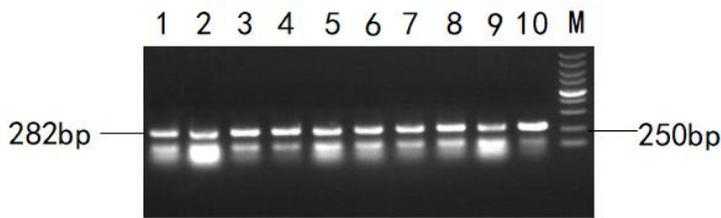


Figure 3 PCR detection of colony of clonal bacteria

Note: M: 5000 bp DNA Marker; 1~10: Clonal bacteria

ProtParam

User-provided sequence:

```

10      20      30      40      50      60
ATGAGTACT ACAATCAACA ACAACCCOCT GTTGGTGTCC CTCOCCGCA AGGATATCCA

70      80      90      100     110     120
COGGAAGGT ACCCTAAGGA CGCATACOCA OCACOGGGT AOCCACAACA GGGTTAOCCA

130     140     150     160     170     180
CCACAGGGT ACCCTCAACA AGGTTACOCT OCTCAAGGGT AOCCTOCCA GTATGGTGCT

190     200     210     220     230     240
CCAACCCCTC AACACAACA GCAACAATCT GGTAGCACTG GCTTCATGGA AGGATGTTTG

250     260     270     280
GCTGCACTGT GCTGTTGCTG ICTOCTAGAT GCATGCTTCT GA

```

Figure 4 Full length of CYS gene

1.3 Analysis of the CYS protein energy domain and hydrophilicity

We used ProtScale software for hydrophilicity analysis and the results showed that the average hydrophilicity range was between -2~2, the hydrophilicity coefficient of the protein is -0.490, indicating that it has some hydrophilic properties (Figure 5).

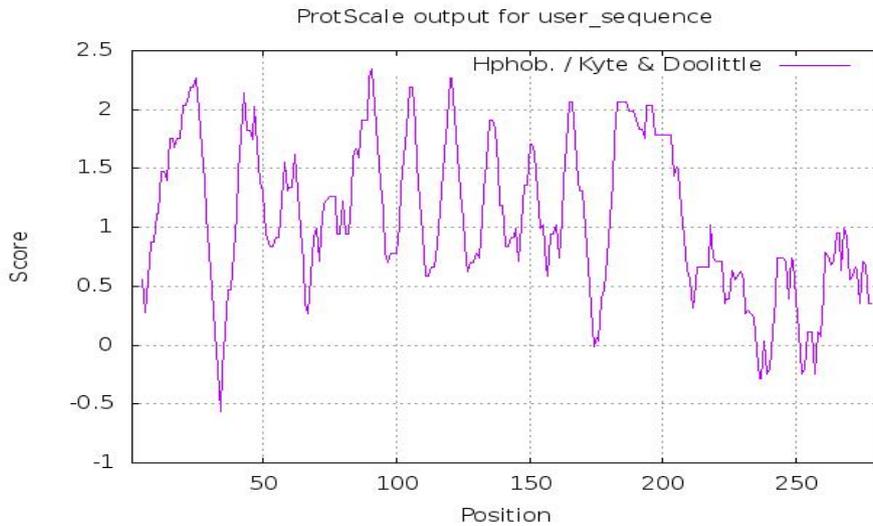


Figure 5 CYS affinity, hydrophobic analysis

1.4 Analysis of the secondary structure of CYS and its protein transmembrane

The secondary structure of the CYS gene was analyzed by importing it into SOPMA software and the main structural forms were found to be α -helix (19.35%) and random coil (65.59%), in addition to extended strand (10.75%) and a β -turn (4.3%) (Figure 6).

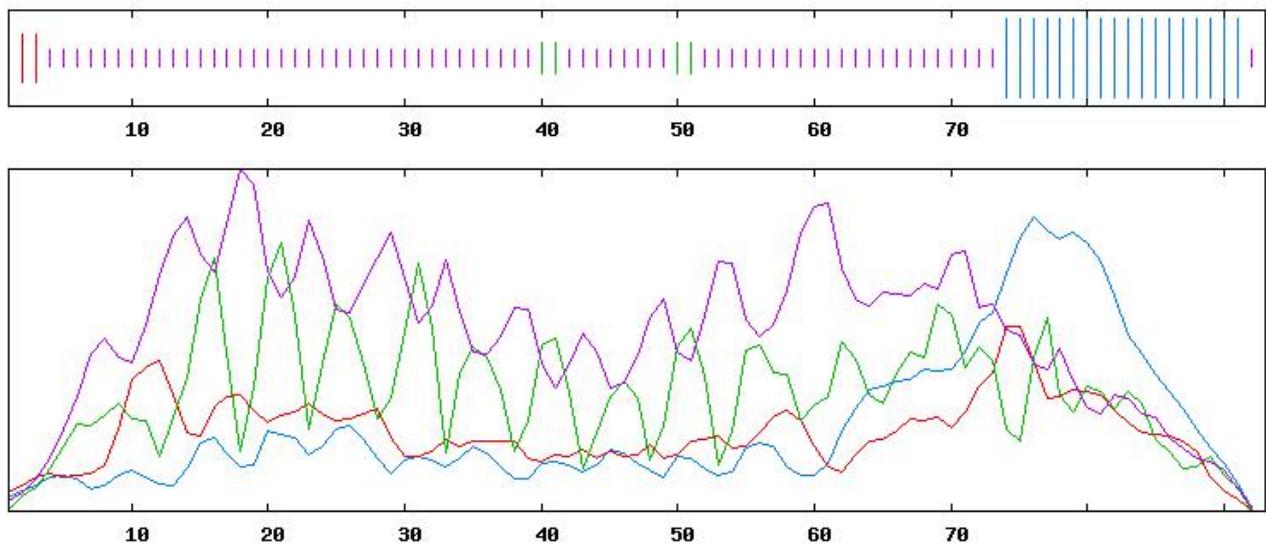


Figure 6 Prediction and Analysis of Secondary structure of BcMYC1

Note: Blue: Alpha helix; Purple: Random coil; Red: Extended strand; Green: Beta turn

Analysis of the transmembrane region of the tobacco CYS gene using TMHMM-2.0 software showed that the number of predicted TMHs for this protein was 0, presumably the protein does not have a transmembrane region and acts outside the membrane (Figure 7).

The SWISS-MODEL database was used to model and predict the homology of the protein encoded by the CYS gene, and a three-dimensional spatial model of the protein was constructed (Figure 8). Meanwhile, a protein with high similarity to it was obtained by comparison in the database: the 6seh.1.B protein. This protein belongs to the structure-specific nucleic acid endonuclease subunit SLX4, with a coverage of 0.20, sequence identity of 21.05 and similarity of 0.32.

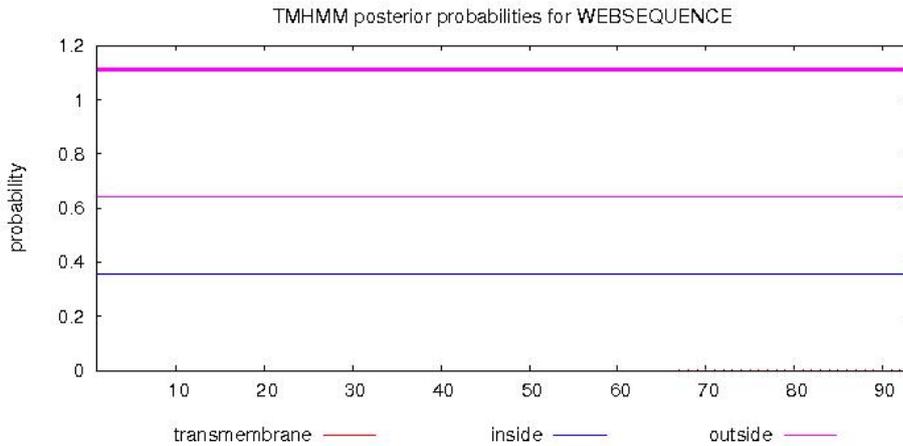


Figure 7 Transmembrane analysis of CYS protein
 Note: Red line: Transmembrane region; Blue line: Inside the membrane; Pink line: Outside the membrane

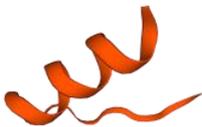


Figure 8 CYS gene homology modeling 3D protein model

1.5 CYS affinity analysis and multiple sequence alignment

A phylogenetic tree was constructed in MEGA4 software comparing the amino acid sequences of the CYS gene with those of other plants (Figure 9), and the results showed that the CYS gene in pepper was the most similar in affinity to the CYS gene in tobacco, followed by HTC and WIH2 in the tomato gene.

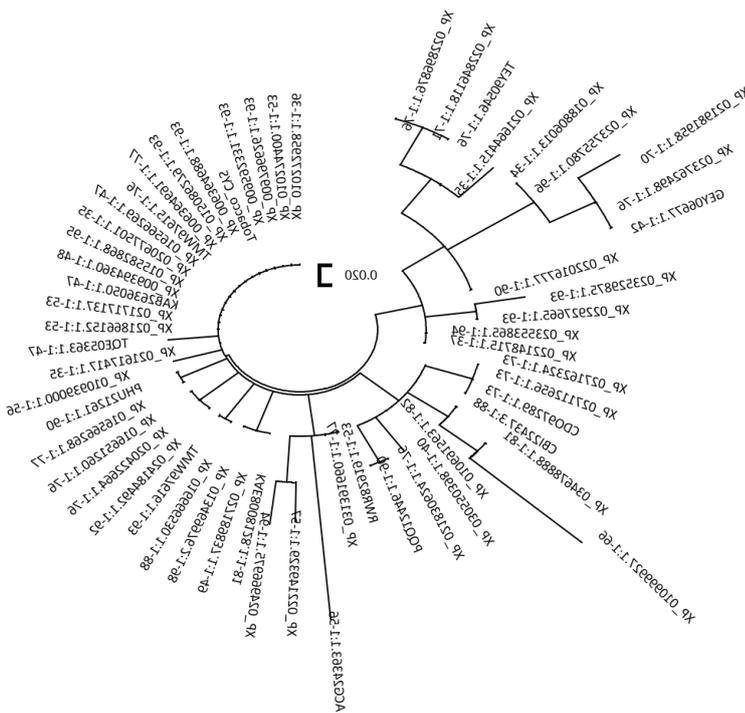


Figure 9 CYS protein phylogenetic tree

2 Discussion

In this study, the CYS gene was cloned from tobacco, and the sequence analysis revealed that the CYS protein is hydrophilic and that the CYS protein does not pass through the trans-membrane region, but acts outside the membrane. The tobacco CYS gene is most closely related to pepper and tomato in the plant.

At present, there are few bioanalyses of cysteine in tobacco, with most focusing on gene and protein sequence analysis of tobacco cysteine proteases and bioinformatics analysis of selenocysteine methyltransferase genes. The structural analysis of the tobacco cysteine protease gene by Yin et al. (2018) showed that the members of NtCP differed between exons and introns and that NtCP had a core conserved motif region (Motif-1-Motif-2-Motif -4), with Motif-1 and Motif-2 both containing glutamine and cysteine, while both NtCP contained signal peptides. And NtCP is a stable protein, providing an important safeguard for its involvement in plant physiological responses (Ma et al., 2018). Tobacco cysteine protease has a trans-membrane structure and is at the N-terminal of the protein sequence, suggesting that NtCP is a trans-membrane protein, which facilitates the N-terminal outside the membrane to receive the signaling molecule and thus act on the C-terminal of the protein sequence (Slee et al., 1999). Gene bioinformatics analysis of selenocysteine methyltransferase (CsSMT) revealed that the product encoded by CsSMT is a stable hydrophilic protein without trans-membrane structure and signal peptide, and was located in the cytoplasmic matrix. And the CsSMT gene mRNA clusters with the *Arabidopsis* homocysteine s-methyltransferase gene (Liu et al., 2013). The above bioinformatics analysis of cysteine proteases and methyltransferases provides a reference for this study and also provides a fuller complement to the bioinformatics aspects of cysteine.

Studies on cysteine in plants have mainly focused on peppers, tomatoes, wheat and other plants, such as the research of cysteine on plant senescence, salt tolerance and resistance to osmotic stress. The cysteine-/histidine-rich DC1 domain protein gene *CaDC1* was found to play a positive regulatory role in plant defense during microbial infection in peppers. Cultivated tomato (*Solanum lycopersicum* Zinc Finger2 [*SIZF2*]) is a cysteine-2/histidine-2 type zinc finger transcription factor with an ERF-associated amphipathic repressor structural domain that is widely expressed during plant development and is capable of delaying senescence and improving salt tolerance in tomato. Structure and function of cysteine enriched protein 19K are encoded by wheat mosaic virus. Mutations in Cys8, Cys11, Cys39, and Cys49 at the N-terminus of 19K-crp were found to greatly reduce the stability of 19K-cys. All these studies were on the role of cysteine for plant development during plant growth, providing a basis for better breeding of plants and an accordance for the development of this study.

In this study, the down-regulation of CYS expression was hypothesized to be most likely associated with changes in plant phenotypes, as understood through bioinformatics analysis of CYS. Trans-membrane proteins are located at the junction between the cell and the outside world, mediating cell-to-external signaling and performing a number of very important cell biological functions (Song and Zhang, 2009). The bioregulation of transmembrane proteins in plants is closely related to other biosynthetic pathways.

3 Materials and Methods

3.1 Test materials

The tobacco material required for this experimental study was sourced from the Northwest Institute of Plateau Biology, Chinese Academy of Sciences. Both purple tobacco transgenic Samsun and green tobacco Samsun were grown in the Plant Culture Room of the Key Laboratory of Molecular Breeding of Wheat in Qinghai Province and were kept and grown in isolation. The growing environment was 25°C with 8 h of light per day.

3.2 Genomic RNA and cDNA preparation

Tobacco RNA extraction. The following centrifugation operations were carried out at 4 °C and at 12 000 r/min. The steps are as follows. (1) Take a medium sized piece of tobacco leaf, add a small amount of solid sodium sulfite, freeze it through liquid nitrogen and grind it to powder and place it in a centrifuge tube. (2) Add 500 µL of TRIzol extract and mix well. (3) After centrifugation for 5 min, the upper liquid layer was separated and 200 µL of trichloromethane was added to it and shaken. (4) Centrifuge for 15 min, then remove the top layer of liquid and

add 0.5 mL of isopropanol dropwise and mix well. (5) After 10 min of centrifugation, the upper layer was poured off and the precipitate was washed with 1 mL of ethanol dropwise. (6) After centrifugation for 5 min, pour out the upper liquid and add 40 μ L of DEPC water to dissolve (Yu et al., 2017).

Tobacco RNA reverse transcription. Specific operations: Add 2 μ L of tobacco RNA to 1 μ L of Oligo(dT)₁₈, 10 μ L of 2 \times TS Reaction Mix, 1 μ L of Trans Script RT respectively, make up to 20 μ L with RNase free water, incubate at 42 °C for 30 min, heat at 85 °C for 5 min, then ice bath for 5 min and centrifuge briefly to obtain tobacco cDNA (Liu et al., 2016).

3.3 Isolation and cloning of the CYS gene

Make the obtained cDNA amplified by using high fidelity DNA polymerase on the GeneAmp PCR System 9 700 PCR instrument along with CDS sequences and the PCR products were examined by agarose gel electrophoresis. The PCR products were then recovered using the SanPrep Column DNA Gel Recovery Kit. 0.2 μ L of *Taq* DNA polymerase (Tiangen) 72 °C for 10 min plus A was added to the resulting object, followed by T-ligation with the pGEM-T Easy cloning vector. The ligated product was transformed into *E. coli* DH5 α receptor state, coated onto ampicillin-resistant LB medium and incubated at a constant temperature of 37 °C for 12~15 h. Positive clones were screened by running gel and punctured and sent to BGI Genomics for sequencing. Positive plasmids were extracted using the SanPrep Column Plasmid DNA small volume extraction kit.

3.4 Analysis of the CYS gene sequence and its encoded protein

In this study, the required gene fragments were selected by Vector NTI Suite 9.0 software, homology matched by the online database of NCBI (www.ncbi.nlm.nih.gov). Then the cDNA sequence open reading frame of CYS gene was analyzed in ExPaSy-Protparam (<http://expasy.org/tools/protparam.html>) software, and its amino acid sequence was speculated. Using the BlastP software (<https://acronyms.thefreedictionary.com/BLASTP>) in the NCBI module and the ProScale software, the conservative region and hydrophobicity of the protein were predicted for the amino acid sequence of the tobacco target gene. The trans-membrane region of its protein sequence was elucidated using TMHMM-2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>). A 3D protein model for homology modelling of the CYS gene was constructed using SWISS-MODEL software (<http://www.swissmodel.expasy.org/interactive>). The amino acid sequences encoded by the tobacco CYS gene were compared with those of other related plants in the MEGA4 online software (www.megasoftware.net/mega4/), and the homology of the amino acid sequences of several organisms with close affinity to the CYS gene was analyzed.

Authors' Contributions

LXY was the experimental designer and executor of the experimental study, completed the data analysis, and wrote the first draft of the paper. GJM and LMH were involved in the design of the experiment and the analysis of the results. LBL was the proposer and person in charge of the project, directing the experimental design and data analysis. WL, ZY and CD completed supervised thesis writing and revision. All authors read and approved the final manuscript.

Acknowledgements

This study was jointly supported by the Applied Basic Research Project of Qinghai Province (2018-ZJ-762) and the Qinghai Province Science and Technology Achievement Transformation Special-Ex-Grant (2018-NK-133).

References

- Liu S.C., Yan D.H., and Wei J., 2013, Bioinformatics analysis of selenocysteine methyltransferase gene in tea, *Xinan Nongye Xuebao* (Southwest China Journal of Agricultural Sciences), 26(6): 2221-2226.
- Liu X.L., Zhu M.Q., Cai W.J., Liu Y.N., Shi Y.N., and Qu C.Q., 2016, Comparison of extraction methods of total RNA from rhizome of *Atractylodes macrocephala*, *Fuyang Shifan Xueyuan Xuebao* (Journal of Fuyang Normal University (Natural Science)), 33(1): 46-49.
- Liu X.Q., Li J, Liu W.J., Zhang Y.L., Zheng T.T., and Wang Y.M., 2014, Research progress on utilization and extraction of tobacco protein, *Huagong Keji* (Science & Technology in Chemical Industry), 22(6): 67-70.
- Ma S.M., Luo J.Z., Wang D.X., Hu Y.X., Fan Z.Y., and Su J.E., 2018, Biological analysis of tobacco cysteine protease, *Anhui Nongye Kexue* (Anhui Agricultural Science), 593(16): 17-19.
- Maciel F.M., Salles C.M.C., Retamal C.A., and Gomeset V.M., 2011, Identification and partial characterization of two cysteine proteases from castor bean leaves (*Ricinus communis* L.) activated by wounding and methyl jasmonate stress, *Acta Physiologiae Plantarum*, 33(5): 1867-1875.

- <https://doi.org/10.1007/s11738-011-0730-z>
Roberts I.N., Caputo C., Criado M.V., and Funk C., 2012, Senescence-associated proteases in plants, *Physiologia Plantarum*, 145(1): 130-139.
- <https://doi.org/10.1111/j.1399-3054.2012.01574.x>
Slee E.A., Harte M.T., Kluck R.M., and Wolf B.B., 1999, Ordering the Cytochrome c-initiated Caspase Cascade: Hierarchical Activation of Caspases-2,-3,-6,-7,-8,and-10 in a Caspase-9-dependent Manner, *Journal of Cell Biology*, 144(2): 281-292.
- <https://doi.org/10.1083/jcb.144.2.281>
Song J.H., and Zhang L.X., 2009, Research progress of plant transmembrane proteins, *Shengwuxue Zazhi (Journal of Biology)*, 26(6): 62-64.
- Wang X.F., Yang L.J., Dong X.N., Li Z.X., and Jiao C.J., 2011, Research progress in the synthesis and regulation of plant cysteine, *Zhiwu Shengli Xuebao (Journal of Plant Physiology)*, 47(1): 37-48.
- Wang Y., Wang X.X., Quan H., Yu X.Y., Liu C., Xiao Q., Li W.F., Xiao Z.L., and Cao Z., 2018, Research and application of L-cysteine sensing detection, *Huaxue Chuanganqi(Chemical Sensor)*, 38(2): 23-33.
- Xiao H.J., Yin Y.X., Chai W.G., and Gong Z.H., 2014, Silencing of the CaCP gene delays salt-and osmotic-induced leaf senescence in *Capsicum annuum L.*, *International Journal of Molecular Sciences*, 15: 8316-8334.
- <https://doi.org/10.3390/ijms15058316>
Yin X.Y., Yang L., Luo W.L., Zhao S., Liu B., and Wang G.Y., 2018, Gene and protein sequence analysis of cysteine protease in tobacco, *Tianjin Nongye Kexue (Tianjin Agricultural Sciences)*, 24(9): 1-4.
- Yu N.T., Zhou Q.L., Luo Z.W., Hu F.C., Zhang Z.L., and Liu Z.X., 2017, Improvement and analysis of cactus total RNA extraction method, *Zhongguo Nongye Kexue (Guangdong Agricultural Sciences)*, 44(3): 75-79.
- Zang Q.W., Wang C.X., Li X.Y., Guo Z.A., Jing R.L., Zhao J., and Chang X.P., 2010, Isolation and characterization of a gene encoding a polyethylene glycol-induced cysteine protease in common wheat, *Journal of Biosciences*, 35(3): 379-388.
- <https://doi.org/10.1007/s12038-010-0043-1>
Zhao J., Ke X., Xu C.H., Li J.Y., and Gong M., 2012, Effects of different light qualities on activity and gene expression of Caspase-like proteases in tobacco leaves, *Agricultural Science & Technology*, 338: 276-279.
- Zong Y., Li S.M., Xi X.Y., Cao D., Wang Z., Wang R., and Liu B.L., 2019a, Comprehensive influences of overexpression of a MYB transcription factor regulating anthocyanin biosynthesis on transcriptome and metabolome of tobacco leaves, *International Journal of Molecular Sciences*, 20(20): 5123.
- <https://doi.org/10.3390/ijms20205123>
Zong Y., Zhu X.B., Liu Z.G., Xi X.Y., Li G.M., Cao D., Wei L., Li J.M., and Liu B.L., 2019b, Functional MYB transcription factor encoding gene AN2 is associated with anthocyanin biosynthesis in *Lycium ruthenicum* Murray, *BMC Plant Biol.*, 19: 169.
- <https://doi.org/10.1186/s12870-019-1752-8>