Gene clone and bioinformatics analysis of subtilisin-like protease in Cucumis sativus

*Xue Wang1, Guangchao Yu1, Xiangyu Wang1, Chunmao Lv2**

[1]College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang 110866, China

[2] College of Food Science, Shenyang Agricultural University, Shenyang 110866, China

Authors' contributions

This work was carried out in collaboration between all authors. CL designed the study. Authors XW analyed the data and wrote the manuscript. Author XW and GY conducted the experiments. All authors read and approved the manuscript.

ABSTRACT

Background: Cucumber target leaf spot (TLS), caused by *Corynespora cassiicola* (*C. Cassiicola*), is a serious disease in cucumber (*Cucumis sativus*) production worldwide. Therefore, cultivating new varieties of TLS resistance of *C. sativus* is an important goal of cucumber breeding. Previous studies have shown that subtilisin-like protease (SUBP) plays an important role in response to *C. Cassiicola* infection in resistant plants.

Objective: In this study, the full-length cDNA of the *CsSUBP* gene was cloned, and the prokaryotic expression vector was successfully constructed in order to study the effects of subtilisin. Futhermore, vital clues regarding *CsSUBP* gene involved in TLS resistance of *C. sativus* are gained from the bioinformatics assay.

Method: The *CsSUBP* gene was identified by sequencing with the intermediate vector pMD18 by designing specific primers and PCR amplification techniques. The prokaryotic expression vector pET30a-*CsSUBP* was further constructed and identified by colony PCR and *EcoR* V and *Sal*I double digestion.

Result: The primary structure of CsSUBP was predicted and analyzed by bioinformatics analysis. The results showed that CsSUBP was weakly acidic protein, N-terminal signal peptide region, including a Inhibitor_I9 domain domain which may be the primary disease resistance domain of the protein.

Conclusion: The pET30a-*CsSUBP* prokaryotic expression vector was constructed successfully. This study is convenient for the study of prokaryotic expression and its kinase activity.

Key words: Subtilisin protease; Gene clone; Bioinformatics assay; Prokaryotic expression

---

*Corresponding author. E-mail:bt_lcm@126.com

1. INTRODUCTION

Cucumber (*Cucumis sativus*) is one of the largest vegetable crops in China. In recent years, cucumber target leaf spot (TLS), caused by *Corynespora cassiicola* (*C. cassiicola*), is a major disease in cucumber production [1]. At present, the prevention and treatment of TLS are mainly chemical control, chemical pesticide can be used to inhibit the activity of pathogens. However, *C. cassiicola* can easily mutate with environmental changes and develop resistance to various chemical pesticide, so it is difficult to prevent and eliminate the disease efficiently [2]. The best method for controlling the disease in cucumber is the use of resistant cultivars, but few high resistant varieties of cucumber can enhance resistance of *C. cassiicola* at present. Therefore, understanding the molecular mechanisms and identifying the molecular components involved in the defense responses of the TLS resistant cucumber cultivars will provide a promising approach to restraining this disease.

In our previous study, a kind of subtilisin-like protease (SUBP) from resistant cucumber plants (Jingyou 38) after the fungus inoculation were explored using iTRAQ quantitative proteomics technique. It is speculated that the SUBP may play an important role in cucumber resistance to early invasion of *C. cassiicola*.

SUBP is a serine protease widely found in plants, bacteria, fungi, and parasites [3-8]. Such proteases have a typical Asp/Ser/His catalytic domain in structure. Studies have shown that it is closely related to the stability of the fungal phenotype and the pathogenicity of the pathogen. It plays an important immune stimulating role in the interaction between plants and pathogens [9]. SUBP is also involved in plant cell programmed cell death (PCD) processes [10]. A number of studies have demonstrated the role of SUBP in plant and pathogen interactions. For example, two subtilisin-like proteins, SAS-1 and SAS-2, in oat (Avena sativa) exhibit specific activity of mammalian caspase apoptosis protease and participate in the signal transduction of PCD during the non-affinity interaction of *oat-Cochliobolus victoriae*, thereby inducing HR production [11-12]. Overexpression of Arabidopsis subtilases gene *AtSBT3.3* activates downstream immune signaling and induces innate immune response in *Arabidopsis* [13].

In this study, the cDNA sequence of *CsSUBP* gene was cloned and the prokaryotic expression vector was constructed. The primary structure of CsSUBP was predicted and analyzed using bioinformatics analysis methods. On this basis, we can further explore the defense function and molecular mechanisms of SUBP against *C. cassiicola* in cucumber.

2. MATERIALS AND METHODS

2.1 Plant Materials

The cucumber varieties used in the experiments were Jinyou 38 (preliminary experiments showed that this variety was resistant to TLS), which were planted in a greenhouse at 28°C under 16:8 light/dark cycles.

2.2 Full-length cDNA cloning

Two candidate genes were screened through preliminary test on the basis of the proteomics analysis. The cDNA sequences of *CsSUBP* (A0A0A0K993) from resistant cucumber leaves were clone. The sequence of *CsSUBP* were aligned with cucumber genome database using the service provided by http://cucurbitgenomics.org/BLAST. The primers (table 1) were designed by using the DNAMAN. The PCR fragments amplified from the cDNAs were cloned into TA vectors pMD18-T and were sequenced.

Table 1. *CsSUBP* gene primer sequences

Gene name	Primer sequence (5’ to 3 ’)
CsSUBP-F	GATATCATGTCCGGCCAATCCACAGCCCTG
CsSUBP-R	GTCGACCTAATGAAGCCTTGCTGCTCCTCCT

2.3 Construction of recombinant prokaryotic expression vector pET30a-*CsSUBP*

Recombinant plasmid pMD18-T-*CsSUBP* and expression vector pET30a were digested by *EcoR* V and *Sal*Irespectively. The double-digested DNA fragment had the same sticky ends as the pET30a vector backbone, and the two fragments were ligated into a complete recombinant plasmid using T4 ligase and ligated overnight at 16°C and then trans-formed into E. coli DH5α competent cells, blue and white screening. Recombinant prokaryotic expression vector was identified by direct colony PCR. Recombinant prokaryotic expression vector pET30a-*CsSUBP* plasmid was extracted and the product was digested by *EcoR V* and *SalI* respectively. It was digested for 3 h at 37°C in a constant temperature metal bath and electrophoresed on a 1.5% agarose gel. Plasmids with positive clones were sent for bioinformatics sequencing.

2.4 Bioinformatics analysis

The specific bioinformatics [14-17] analysis contents and tools used are as follows see table 2.

Table 2 Bioinformatics analysis content and Tools website

Analyze the content	Name of software	Bioinformatics Analysis Tools Website
Primary structure	ProtParam	http://web.expasy.org/protparam/
Secondary structure	SOPMA	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
Hydrophobiciy	ProtScale	http://web.expasy.org/protscale/
Signal peptide	SignalP 4.0	http://www.cbs.dtu.dk/services/SignalP/
Functional domains	SMART	http://smart.embl-heidelberg.de/

3. RESULTS

3.1 Amplification of *CsSUBP*

The reverse transcription of cDNA was used as the template. PCR amplification product was subjected to 1.5% agarose gel electrophoresis and a fragment of approximately 282 bp was obtained as expected (Fig.1)

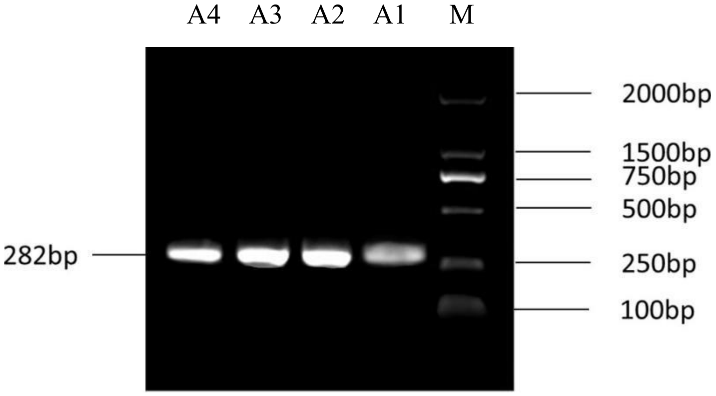


Fig. 1 Electrophoresis profile of PCR products of CsSUBP Gene. A1-A4: *CsSUBP* PCR amplification products; M: DL2000 DNA Marker

3.2 Construction of Prokaryotic Expression Vector

To construct the prokaryotic expression vector, we recycle the small fragments of pMD18-T-*CsSUBP* (Fig. 2, 3) and large fragments of pET30a plasmid after double digest with *EcoR* V and *Sal*I. The colony PCR products of recombinant plasmid pET30a-*CsSUBP* via agarose gel electrophoresis analysis showed 282 bp of the target gene band (Fig.4, 5). Sequencing results showed that the gene fragments was consistent with the target gene sequence in GenBank (Fig. 6).

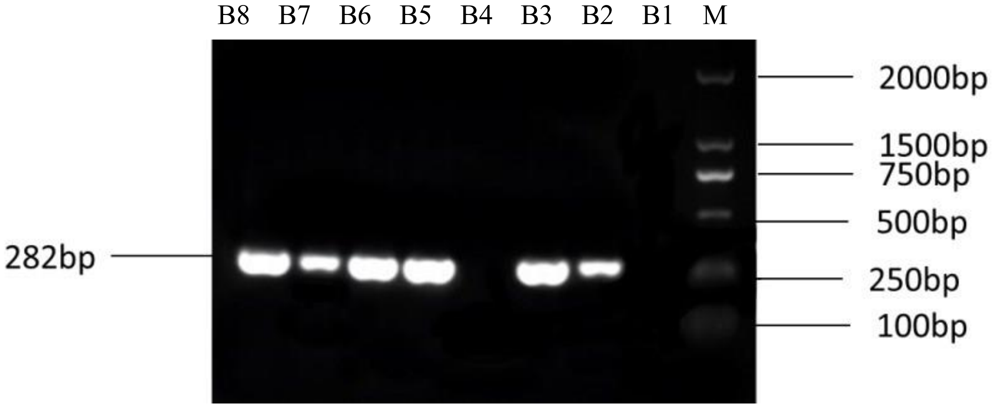


Fig. 2 The Colony PCR of recombinant plasmid pMD18-CsSUBP. B1-B3,B5-B8: Colony PCR of *CsSUBP* gene; B4: water; M: DNA marker DL2000

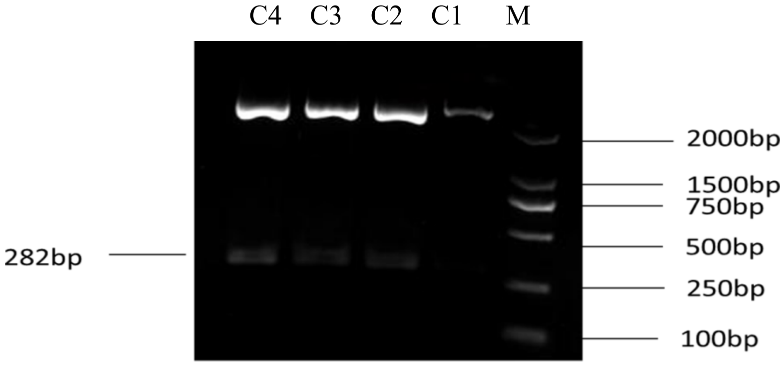


Fig. 3Double digestion of recombined vector pMD18-*CsSUBP* with restriction enzymes.C1-C4: Double digestion products; M: DNA marker DL2000

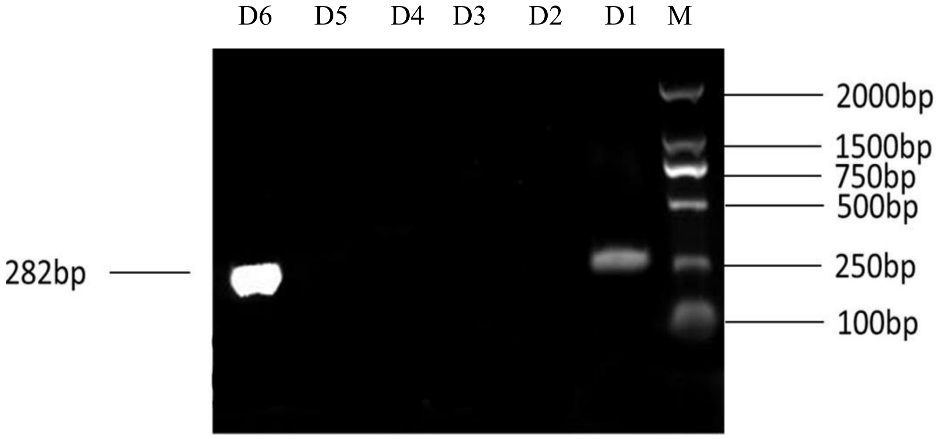


Fig. 4 The Colony PCR of recombinant plasmid of pET30a-*CsSUBP*. D1-D3, D6: Colony PCR products of *CsSUBP* gene; D4, D5: Water control; M: DNA marker DL200

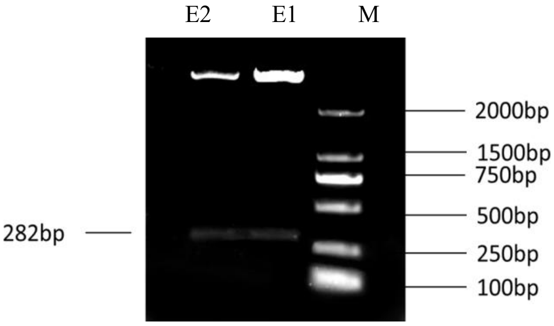


Fig. 5 Double digestion of recombined vector pET30a-*CsSUBP* with restriction enzymes. E1, E2: Double digestion products; M: DNA marker DL2000

```
SUBP.seq                ....................................................................................................      0
subp-6-M13F_-47__B04.ab1RC  GCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGAATTCGAGC    800
Consensus

SUBP.seq                .................................ATGTCCGGCCAATCCACAGCGGTGCACATCGTTTACACTGAGCGGCCTCACAATGAAGAACCCGAGG     67
subp-6-M13F_-47__B04.ab1RC  TCGGTACCCGGGGATCCTCTAGAGATTGATATCATGTCCGGCCAATCCACAGCGGTGCACATCGTTTACACTGAGCGGCCTCACAATGAAGAACCCGAGG    900
Consensus                                                atgtccggccaatccacagcggtgcacatcgtttacactgagcggcctcacaatgaagaacccgagg

SUBP.seq                CCTATCACATCCGAACCCTAGCTTCTGTTCTTGGCAGTGAAGGCGCTGCTAGGGAGGCGCTTGTGTATAGCTATAAAAATGCTGCGAGTGGCTTCTCTGC    167
subp-6-M13F_-47__B04.ab1RC  CCTATCACATCCGAACCCTAGCTTCTGTTCTTGGCAGTGAAGGCGCTGCTAGGGAGGCGCTTGTGTATAGCTATAAAAATGCTGCGAGTGGCTTCTCTGC   1000
Consensus               cctatcacatccgaaccctagcttctgttcttggcagtgaaggcgctgctagggaggcgcttgtgtatagctataaaaatgctgcgagtggcttctctgc

SUBP.seq                TCGGCTTACGTCCGATCAGGTTGCGGAGATTACCAAACAACCAGGAGTATTGCAGGTTGTCCAAAGCGGTAATAATAAGCTGCATTCTGGAGGAGGAGCA    267
subp-6-M13F_-47__B04.ab1RC  TCGGCTTACGTCCGATCAGGTTGCGGAGATTACCAAACAACCAGGAGTATTGCAGGTTGTCCAAAGCGGTAATAATAAGCTGCATTCTGGAGGAGGAGCA   1100
Consensus               tcggcttacgtccgatcaggttgcggagattaccaaacaaccaggagtattgcaggttgtccaaagcggtaataataagctgcattctggaggaggagca

SUBP.seq                GCAAGGCTTCATTAG....................................................                                     282
subp-6-M13F_-47__B04.ab1RC  GCAAGGCTTCATTAGGTCGACAATCGTCGACCTGCAGGCATGCAAGCTGGCACTGGCCGTCGTACTC                                    1167
Consensus               gcaaggcttcattag
```

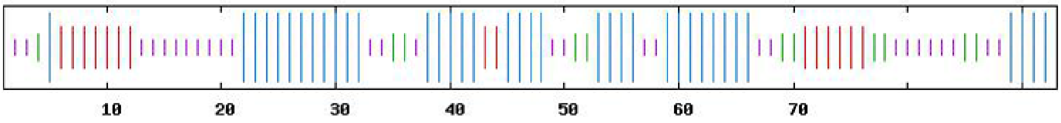
Fig.6 Sequencing results recombinant vector pET30a-*CsSUBP*.

3.3 Bioinformatics Analysis of CsSUBP proteins

Some information about the CsSUBP are obtained by some online tools (shown in table2). The protein molecular weight is 23146.02, chemical formula is $C_{834}H_{1388}N_{282}O_{348}S_{69}$, isoelectric point (pI) is 5.28. It is weakly acidic and hydrophobic average coefficient is 0.794. It is speculated that CsSUBP maybe a hydrophobic protein (shown in Fig 8). Neither signal peptide nor transmembrane region is found in the CsSUBP (shown in Fig 9).

Secondary structure of CsSUBP is analysed by SOPMA. The result showed that CsSUBP consists of 39.78% alpha helix (h), 16.13% extended strand (e), 11.83% beta turn (t) and 32.26% random coil (c)(Fig.7 and Table 3).

```
         10        20        30        40        50        60        70
          |         |         |         |         |         |         |
MSGQSTAVHIVYTERPHNEEPEAYHIRTLASVLGSEGAAREALVYSYKNAASGFSARLTSDQVAEITKQPcccct
heeeeeeeccccccccchhhhhhhhhhhccttchhhhheehhhhcctthhhhcchhhhhhhhcctGVLQVVQSGN
NKLHSGGGAARLHteeeeeettcccccccttcchhhh
```



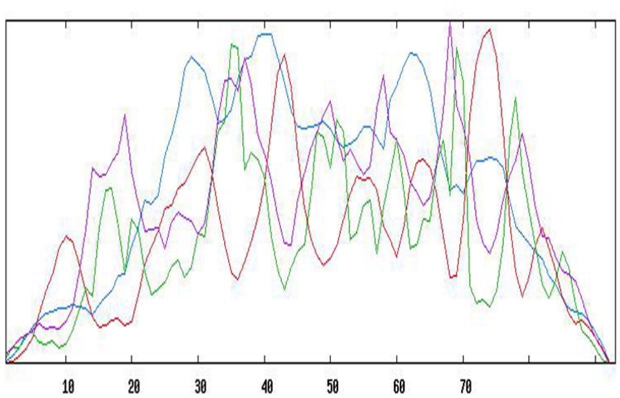


Fig. 7 The secondary structure of CsSUBP protein. h: Alpha helix; t: Beta turn; e: Extended strand; c: Random coil

Table 3 Prediction the secondary structure of CsSUBP protein

Name	Number	Percentage
Alpha helix	37	39.78%
3_{10} helix	0	0.00%
Pi helix	0	0.00%
Beta bridge	0	0.00%
Extended strand	15	16.13%
Beta turn	11	11.83%
Bend region	0	0.00%
Random coil	30	32.26%
Ambigous states	0	0.00%
Other states	0	0.00%

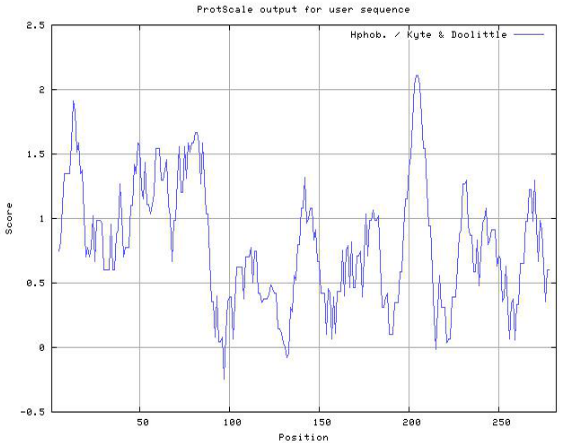


Fig. 8 Prediction the hydrophobicity of CsSUBP protein

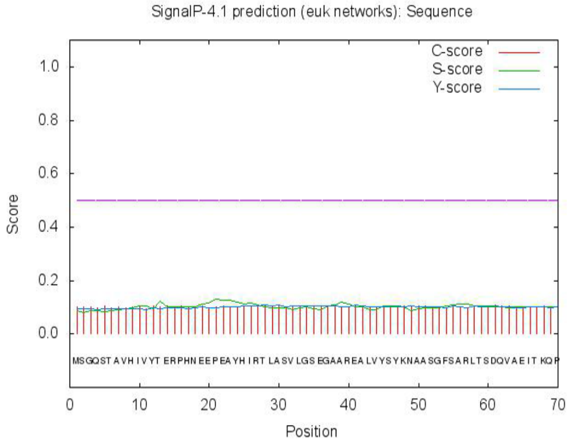


Fig. 9 Prediction the signal peptide of CsSUBP protein

A major functional domains of *CsSUBP*-encoded proteins is found by SMART analysis. The results are shown in Table 4.

Table 4 Functional domain analysis of *CsSUBP*-encoded proteins

Name	Position	E-value
Inhibitor_I9 domain	8(Starting site)-84(Termination site)	4.8e-15

Inhibitor_I9 domain, which belongs to a member of the MEROPS family of protease inhibitors I9, a protease pro-peptide inhibitor (sometimes referred to as an activated peptide) that regulates the folding and activity of the proenzyme or zymogen. The anterior segment extends into the enzyme moiety and shields the substrate binding site, thereby promotes enzyme inhibition. Although it often has lower sequence identity, several such propeptides have similar topologies. The propeptide region has an open-loop, sandwich-like anti-parallel α/ β, it also has two α-helices and four β-strands with a 2x (β/α/β) topology.

CsSUBP also belongs to the N-terminal propeptide domain of the meropsis family S8A, a subtilisin peptidase. The use of the subtilisin propeptide as a chaperone helps to fold the mature peptidase. The propeptide is removed by proteolytic cleavage to remove the activating enzyme.

4. DISCUSSION

Cucumber is one of the important vegetables in the world, and it is also the largest greenhouse productin China. Cucumber has a cool taste and medicinal value. However, in the process of production, cucumber will be exposed to a variety of diseases such as blight, powdery mildew, and target leaf spot. Among them, the incidence of target leaf spot is extremely high, which causing serious economic losses [18]. Studies have shown that breeding and using disease-resistant varieties are the most direct, effective, and economical measures for disease control.

Our iTRAQ analysis on the inoculation of cucumber leaves showed the expression level of a subtilisin-like protease (A0A0A0K993) was significantly decreased at 6 h after pathogen infection. In addition, quantitative real-time PCR technology found that its corresponding genes expression were also decreased a downward trend after inoculation of pathogen. We speculated that it may be involved in the early invasion process of cucumber against *C. cassiicola.*

Subtilisin-like protease (SUBP) is a serine protease, which can regulate plant embryogenesis and organ development, fruit ripening and senescence [19-22]. It also plays an important role in the interaction between plant and pathogen [9]. Plant-pathogen interaction begins with the pathogen contact host and ends with a significant disease or susceptibility response [23]. When plant was attacked by pathogen, it will produce two immune defense lines (MTI and ETI). Studies have shown that in plants such as tomato [24], soybean [25], and wheat [26], the subtilisin-like protease can regulate its immune response and participate in intercellular signal transduction during interaction. However, whether it will stimulate the immune response in cucumber is still unclear.

In this study, bioinformatics analysis of the subtilisin-like protease showed that it contained a propeptide domain (Inhibitor_I9 domain), which may be the main disease resistance domain of the protein.

E. coli prokaryotic expression system has the characteristics of simple operation, short culture period and high expression level, which is beneficial to the study of protein structure and function [27]. In addition, the prokaryotic expression vector pET-30a has the advantages of strict regulation and exquisite design of multiple cloning sites. In this research, the *SUBP* gene was cloned from Jinyou 38 and the prokaryotic expression vector pET30a-*CsSUBP* was constructed to facilitate the establishment of Cs*SUBP*-encoded recombinant protein. It also lays the foundation for studying the role of protein function and kinase activity in plant disease resistance.

5. CONCLUSION

We have successfully amplified and purified of *CsSUBP* gene in cucumber coding sequence, and constructed the recombinant prokaryotic expression vector pET30a-*CsSUBP*. Bioinformatics analysis shows that CsSUBP belongs to the hydrophobic protein and has no signal peptide at the N-terminus. It contains a functional domain. Its protein chemical formula is $C_{834}H_{1388}N_{282}O_{348}S_{69}$, an isoelectric point (PI) is 5.28 and a molecular weight is 23146.02. Our study is a preparation for revealing the functions and mechanisms of CsSUBP against *C. cassiicola* in cucumber.

ACKNOWLEDGEMENTS

This work was supported by Liaoning Provincial Natural Science Fund Guidance Program 20180550919.

DECLARATION OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

1. Hasama W, Morita S, Kato T. Seed transmission of Corynespora melonis, causal fungus of target leaf spot, on cucumber . Jpn.j.phytopathol. 2009; 59(2):175-179.

2. Qi Y X, Zhang X, Pu J J, et al. Morphological and molecular analysis of genetic variability within isolates of *Corynespora cassiicola* from different hosts . European Journal of Plant Pathology. 2011; 130(1):83-95.

3. Siezen R J, De W V, Leunissen J A, et al. Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases . Protein Engineering.1991;4(7): 719.

4.Ksiazek M, Karim A Y, Bryzek D, et al. Mirolase, a novel subtilisin-like serine protease from the periodontopathogen *Tannerella forsythia*. Biological Chemistry. 2015; 396(3): 261-275.

5. Duan X, Zhang Z, Wang J, et al. Characterization of a novel cotton subtilase gene *GbSBT1* in response to extracellular stimulations and its role in Verticillium resistance . Plos One. 2016; 11(4): e0153988.

6. Meyer M, Huttenlocher F, Cedzich A, et al. The subtilisin-like protease SBT3 contributes to insect resistance in tomato . Journal of Experimental Botany. 2016;67(14): 4325-4338.

7. Mizutani O, Shiina M, Yoshimi A, et al. Substantial decrease in cell wall α-1, 3-glucan caused by disruption of the kexB gene encoding a subtilisin-like processing protease in Aspergillus oryza e. Bioscience, biotechnology, and biochemistry. 2016; 80(9): 1781-1791.

8. Pepe A, Frey M E, Munoz F, et al. Fibrin (ogen) olytic and antiplatelet activities of a subtilisin-like protease from *Solanum tuberosum* (StSBTc-3). Biochimie.2016;125: 163-170.

9. Figueiredo A, Monteiro F, Sebastiana M. Subtilisin-like proteases in plant-pathogen recognition and immune priming: a perspective . Front Plant Sci.2014;5(739): 739.

10. Vartapetian A B, Tuzhikov A I, Chichkova N V, et al. A plant alternative to animal caspases: subtilisin-like proteases. Cell death and differentiation. 2011; 18(8): 1289.

11. Rantong G, Evans R, Gunawardena A H. Lace plant ethylene receptors, AmERS1a and AmERS1c, regulate ethylene-induced programmed cell death during leaf morphogenesis . Plant molecular biology. 2015; 89(3): 215-227.

12. Coffeen W C, Wolpert T J. Purification and characterization of serine proteases that exhibit caspase-like activity and are associated with programmed cell death in *Avena sativa*. The Plant Cell. 2004; 16(4): 857-873.

13. Ramírez V, López A, Mauch-Mani B, et al. An extracellular subtilase switch for immune priming in *Arabidopsis* . PLoS Pathogens. 2013;9(6): e1003445.

14. Zhang B, Pan X, Cannon C H, et al. Conservation and divergence of plant microRNA genes. The Plant Journal. 2006; 46(2): 243-259.

15. Zhang B, Wang Q, Wang K, et al. Identification of cotton microRNAs and their targets. Gene. 2007; 397(1): 26-37.

16. Yin Z, Li C, Han X, et al. Identification of conserved microRNAs and their target genes in tomato(*Lycopersicon esculentum*) . Gene. 2008; 414(1): 60-66.

17. Sunkar R, Jagadeeswaran G. In silico identification of conserved microRNAs in large number of diverse plant species . BMC Plant Biology. 2008; 8(1):37.

18.Yang M. Study on the Diversity of Pathogens of Coryneform Leaf Spot in Vegetables in China. Chinese Academy of Agricultural Sciences. 2013.

19.Tanaka H, Onouchi H, Kondo M, et al. A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. Development.2001;128(23): 4681-4689.

20. D' Erfurth I, Signor C.L, Aubert G. A role for an endosperm-localized subtilase in the control of seed size in legumes.

New Phytol.2012;196(3): 738-751.

21. Othman R, Nuraziyan A. Fruit-specific expression of papaya subtilase gene. Plant Physiol.2010;167(2): 131-137.

22. Roberts I.N, Caputo C, Kade M, et al. Subtilisin-like serine proteases involved in N remobilization during grain filling in wheat.Acta Physiol. Plant.2011;33(5): 1997-2001.

23. Hong L I , Zhang Z . Systems understanding of plant–pathogen interactions through genome-wide protein–protein interaction networks. Frontiers of Agricultural Science and Engineering. 2016;3(2).

24. Tornero P, Conejero V, and Vera P. Identification of a new pathogen-induced member of the subtilisin-like processing protease family from plants. Biol. Chem.1997; 272(22):14412-14419.

25. Yamaguchi Y, Barona G, Ryan C.A, et al. GmPep914, an eight-amino acid peptide isolated from soybean leaves, activates defense-related genes. Plant Physiol. 2011;156(2): 932-942.

26. Fan T, Bykova N.V, Rampitsch C, et al. Identification and characterization of a serine protease from wheat leaves. Eur. Plant Pathol. 2016; 146(2): 293-304.

27.Rong J J, Diao Z Y, Zhou G H. Advances in the development of the *E. coli*. Drug biotechnology. 2005;12(6):416-420.