# CLONING AND SEQUENCE ANALYSIS OF PATHOGENESIS-RELATED PROTEIN 1 GENE FROM ANTHRACNOSE IN TEA PLANT

Yun-Long Shi[1], Long-Jie Zhang[2], Kai-Rong Wang[2], Yue-Rong Liang[1]*

[1]Tea Research Institute, Zhejiang University, #866 Yuhangtang Road, Hangzhou 310058, China

[2]Ningbo Huangjinyun Tea Science and Technology Co. Ltd., Yuyao 315412, China

**Abstract**

*Anthracnose of tea plant (Camellia sinensis (L.) O. Kuntze), caused by the fungus Colletotrichum genus, is the major leaf disease in China. To verify the gene molecular characterization, the pathogenesis-related protein 1 (CsPR1) gene was cloned from two tea cultivars. The full length cDNA of CsPR1 consists of 483 bp open reading frame (ORF) encoding a polypeptide with 160 amino acids. Sequence alignment against National Center for Biotechnology Information Search database (NCBI) showed that CsPR1 shared 99.38%, 71.43%, 67.50% homology identity with Camellia sinensis reference genome 'shuchazao', Vitis vinifera and Quercus suber, respectively. It contains a dominate domain superfamily of CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins) domain, which accumulates after pathogen infection. The qRT-PCR results revealed that the gene expression level of CsPR1 was remarkably increased in infected groups compared to control. The resistance cultivar accumulated a higher level protein encoded by CsPR1 than that of susceptible cultivars. These results suggest that CsPR1 plays a critical role in the interaction between host tea plant and Colletotrichum, as well as acts as a signal or a defense marker associated with anthracnose resistance in tea plant.*

**Keywords:** *Anthracnose, CsPR1, resistance, molecular breeding, Camellia sinensis*

## 1. INTRODUCTION

In case of pathogen infection in host plant, a hypersensitive reaction (HR) occurred accompaning with a great increase in defense proteins, e.g., Pathogenesis-related (PR) proteins [1]. According to nomenclature suggested by Van Loon et al., several types of PR proteins comprise 17 family menbers with proteins ID from PR1 to PR17 [2]. The dash (-) between PR and number was omitted for uniformity. These proteins are small, secreted or extracellular, acidic or vacuole-targeted proteins with antimicrobial and antifungal activities through contact toxicity and hydrolytic activities on cell walls *in vitro*, encoded by PR genes induced after pathogen infection [3, 4]. The PR1 protein, which is a dominant group of PR proteins, structurally has four α-helices, four β-strands and some conserved residues with six cysteines [2]. The induction of PR1 protein, whose molecular weight ranging from 14 to 17 kDA, can be found in many plants by pathogens and salicylic acid, acting as a signal of system acquired resistance (SAR) and a marker for salicylic acid-mediated disease resistance [5, 6]. In general, SAR was defined as a broad-spectrum plant disease resistance that induced after pathogen infections [3]. In *Arabidopsis*, salicylic acid (SA) induced the expression of PR1, PR2, and PR5 strongly [7]. *PR1* gene plays an important role that mediates gene induction and plant immunization. PR-1 is the most abundantly produced proteins in plants on pathogen attack, may acting as anti-fungal agents or be involved in cell wall loosening (NCBI Proteins entry cl00133). Breen et al. reported that PR-1 proteins, which targeted by plant pathogens, have sterol-binding activity and an embedded defence signalling peptide [6]. However, the fuction and role of *PR1* in plant defence remains poorly understood though. Therefore, the study in raltion to *PR1* needs urgent and it will be helpful for developing resistant crops and plants.

Anthracnose is one of the most severe foliar fungal disease on tea plant. The yield loss induced by anthracnose was estimated to range from 5% to 20% or even more, which exceeded to 50% in Fuding country in 1992 [8]. *Colletotrichum* species are fungal pathogens that devastate tea plants described as causing anthracnose diseases in China. Many species could survive well on aging leaves formed as dense tiny black dots called sporocarp on the lesion in winter, which would be spread via water-splash

dispersal of conidia [9]. Wounds after harvesting and cutting by machines, as well as hot and humid conditions during raining season in June could facilitate disease transmission synergistically [10, 11]. Fungicides were still used to control anthracnose. To reduce the use of fungicides in tea field, investigating physiological properties of resistance genes and proteins is important for understanding disease control mechanisms and develop new disease control methods.

In the present study, a cDNA sequence of *CsPR1* gene was cloned from anthracnose infected tea leaves and its its sequence, structure, phylogeny was investigated. Its expression pattern and expression levels in healthy and desease infected leaves throughout two cultivars were also analyezd. The study will provide useful information for be breeding resistant cultivars with tea plant immunity against pathogens using molecular marker assistant selection and molecular breeding methods.

## 2. MATERIALS AND METHODS

### *2.1. Plant Materials*

Two tea cultivars 'Fudingdabaicha' (FD) and 'Longjing43' (LJ43) were used in this study. FD is a well-known resistant tea cultivar to anthracnose and LJ43 is susceptible one, both of which were grown at the Experimental Tea Farm of Zhejiang University, Hangzhou, China. For each cultivar, ten healthy leaves and ten anthracnose infected leaves after petiole and speckle removal, were collected in tin foil paper, frozen in liquid nitrogen immediately and stored in a refrigerator at -80 °C, respectively.

### *2.2. RNA extraction and first-strand cDNA generation*

Total RNA was extracted from healthy and infected leaves of FD andLJ43, using RNAprep Pure Kit for Polysaccharides & Polyphenolics rich plant (Tiangen Biotech Co., Ltd., China) according to the manufacturer's instruction. The integrity and concentration of RNA were measured by agarose gel electrophoresis and Agilent 2100 Bioanalyzer (Agilent Technologies Inc., USA). First strand cDNA was generated using PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa, Japan) and stored at at -20 °C for further analysis.

| Primer name | Primer sequence(5'-3') |
|---|---|
| **total*CsPR1*-F** | GGAAATCAAAATGGAGCTTTC |
| **total*CsPR1*-R** | CAAAAACCACCCTAGTATGG |
| **q*CsPR1*-F** | TGTCGGATGCAGTGAAGATG |
| **q*CsPR1*-R** | ACCCTATTTGTGCGGTGTT |
| ***Actin*-F** | CTTCCTCATGCTATCCTCCGTCTT |
| ***Actin*-R** | ATTTCCCGTTCAGCAGTGGTG |

**Table 1.** Primer sequences for *CsPR1* gene

### *2.3. Cloning and sequencing of CsPR1 gene*

The full-length *CsPR1* gene was amplified by RT-PCR (Reverse Transcription-Polymerase Chain Reaction) using primers (total*CsPR1*) in Table 1. The primers were designed according to previous transcriptome databases using PrimerSelect in DNAStar package (DNAStar Inc, USA). The RT-PCR reactions were performed using the following conditions: reaction system comprising 2 μL cDNA, 0.5 μL primers each (20 Mm), 25 μL PrimeSTAR Max DNA Polymerase (TaKaRa, Japan), 22 μL dd $H_2O$, 98 °C for 2 min, followed by 35 cycles of 98 °C for 10 s, 55 °C for 15 s, and 72 °C for 20 s, and a final extension for 10 min at 72 °C. The products were fractionated by electrophoresis on 1.2% (w/v) agarose

gel and extracted by MiniBSET Agrose Gel DNA Extraction Kit (TaKaRa, Japan). The obtained amplicons were cloned into pMD18-T vector, transformed into *E.coli* and sequenced by Sangon Biotech (Shanghai, China) Co., Ltd.

*2.4. Sequence analysis*

The nucleotide sequence of *CsPR1* was identified using BLAST tools at National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/). The deduced amino acid sequence analysis and sequence alignment was performed using the software DNASTAR's Lasergene package (DNASTAR Inc, USA) and DNAMAN (Lynnon Corporation, Canada). The phylogenetic tree was constructed by software MEGA7 [12] with high identity sequences over 20 species. The physicochemical properties of proteins encoded by *CsPR1* were analyzed using ProtParam tools (http://expasy.org/tools), TargetP 1.1 Server and SingalP-5.0 Server (http://www.cbs.dtu.dk/services). The secondary and tertiary structure were predicted using GOR IV (https://npsa-prabi.ibcp.fr) and I-TASSER (Iterative Threading ASSEmbly Refinement) [13].

*2.5. Gene expression analysis*

To detect the expression of *CsPR1* in FD and LJ43, quantitative real-time PCR (qRT-PCR) was performed using primers (q*CsPR1*) in Table 1 with *Actin* as internal control StepOnePlus Real-Time PCR System (Applied Biosystems, USA) by using SYBR Premix Ex TaqTM II (TaKaRa, Japan), based on the previous protocols [14]. The triplicate qRT-PCR analysis was calculated using the $2^{-\Delta\Delta CT}$ method [15] with the expression values normalized against that of FD_H. Relative gene expression data was expressed as mean ± SD.

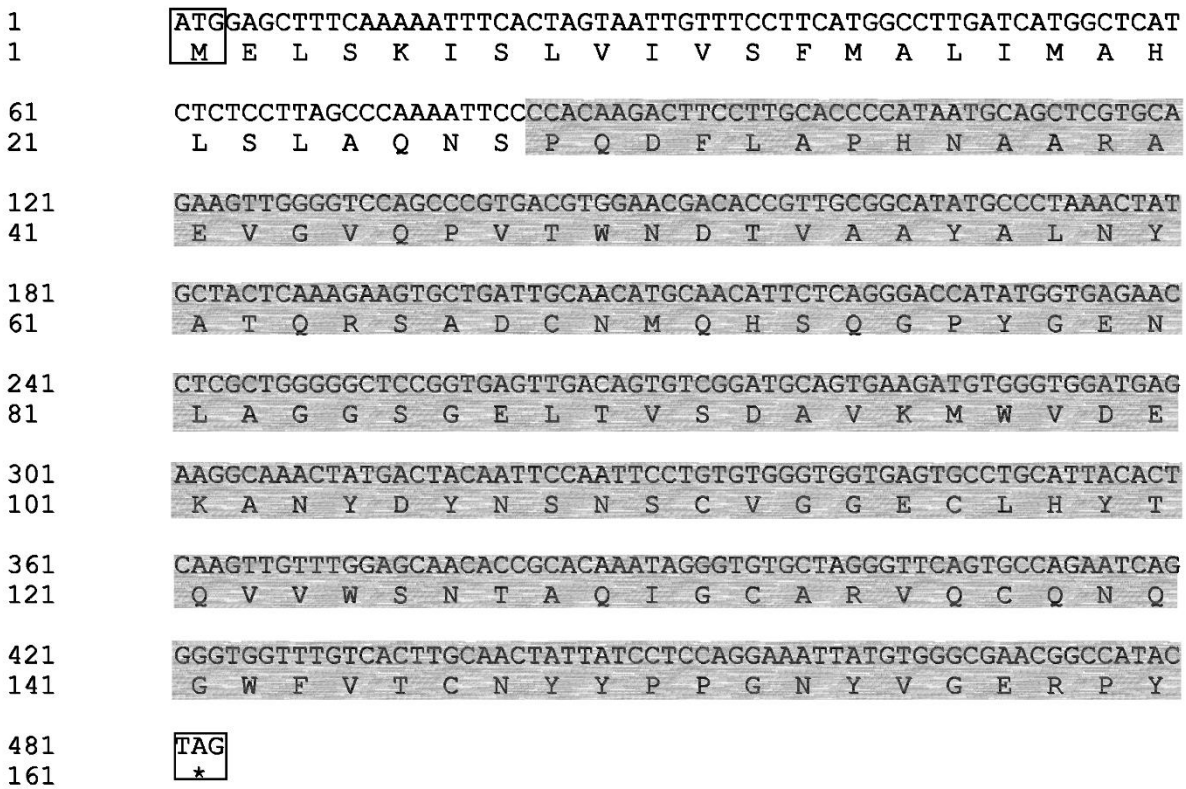



**Fig. 1.** Nucleotide sequence and predicted amino acids sequence of *CsPR1*. Black boxes indicate the start and stop codons. The grey region indicates the CAP domain.

## 3. RESULTS AND DISCUSSION

### *3.1. CsPR1 sequence analysis*

Using the RNA products, a specific primer pair (Table 1, total*CsPR1*-F/totalCsPR1-R) and RT-PCR, the *CsPR1* gene was cloned and sequenced. The amplified product consists of 483 bp open reading frame (ORF), which encodes a protein with 160 amino acids (Fig.1). The sequeces of *CsPR1* in LJ43 and FD are all the same and have been deposited in GenBank under accession number MH454566.1(LJ43) and MH454567.1(FD). Specifically, the deduced protein contained the the CAP domain (grey region in Fig.1), which comprised 137 amino acids in line with approximately 150 amino acids [6]. CAP family proteins are secreted glycoproteins, highly stable in the extracellular fluid, implicated and conseved in immune defense in bacteria, fungi, animals and plants over 2500 species [6, 16]. Genetic and biochemical evidence suggested that PR1 and CAP domains served to inhibit pathogen growth through binding, sequestrating and extracting sterols from the pathogen membrane [16, 17].

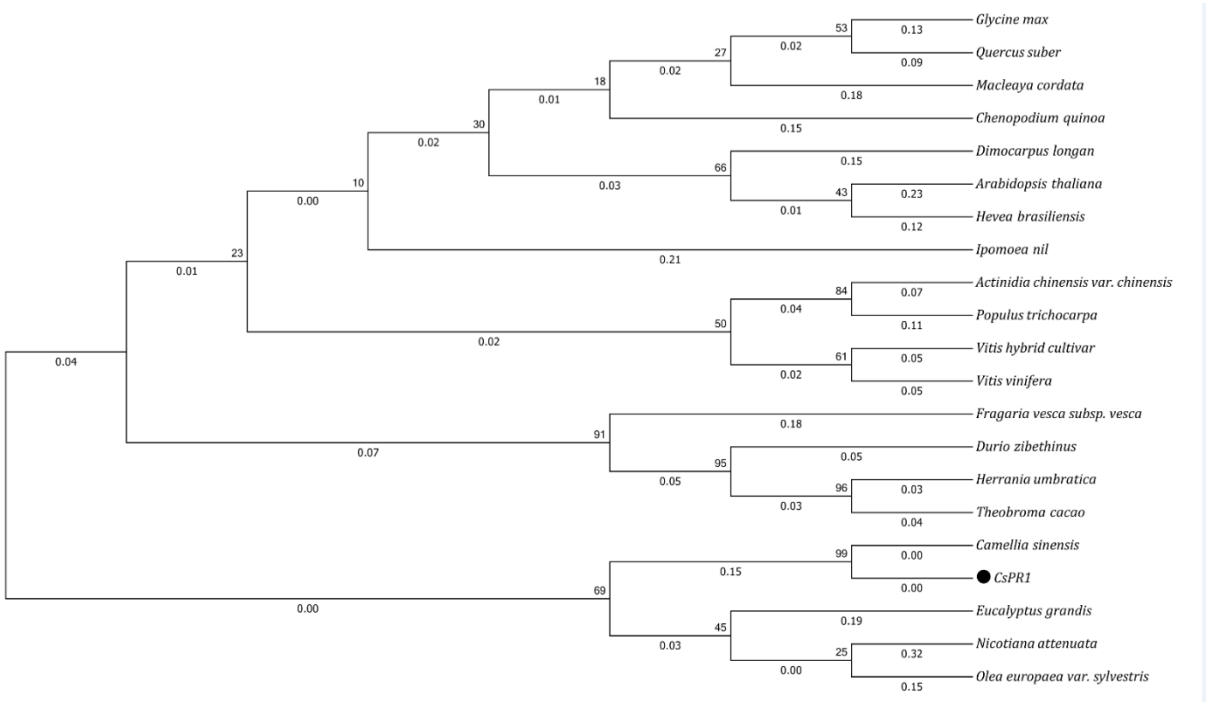



**Fig. 2.** Phylogenetic tree of homologous pathogenesis-related 1 proteins. Sequences are presented by species name. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches [18]. The evolutionary distances were computed using the Poisson correction method, whose values are shown in the middle of the line.

Homology analysis aginst NCBI non-redundant protein sequences database using BLAST showed that CsPR1 protein shared 99.38% identity with pathogenesis-related protein 1-like of *Camellia sinensis* reference genome 'shuchazao' (XP_028054681.1), 71.43% identity with that of *Vitis vinifera* (XP_002274307.1), 67.50% identity with that of *Quercus suber* (XP_023905814.1), 65.64% identity with that of *Olea europaea* var. *sylvestris* (XP_022885095.1) and 59.48% identity with that of *Arabidopsis thaliana* (CAA65420.1). The homologs found in the BLAST database were used to construct a phylogenetic tree and sequences were selected to include available plant species as many as possible and to avoid redundancies. As shown in Fig. 2, our results revealed that these plant PR1 proteins were classified into two major branches and CsPR1 located in the underlying branch, indicating that CsPR1 might had a similar protein structure, function and evolutionary relationship to that of *Eucalyptus grandis*, *Nicotiana attenuata*, *Olea europaea* var. *sylvestris*. Based on sequence alignment with PR1 protein of *Camellia sinensis* and *Arabidopsis thaliana*, PR1, especially CAP domain were highly conserved. The consensus of amino acids was indicated in Fig.3.

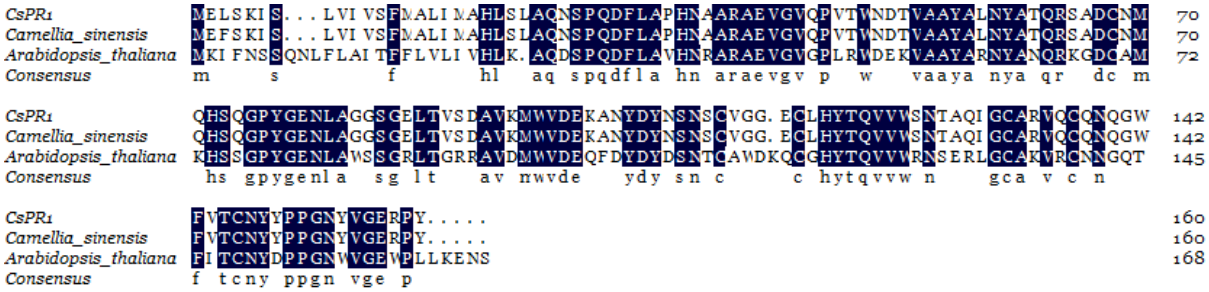



**Fig. 3.** Sequence alignment of CsPR1 protein against *Camellia sinensis* (reference genome 'shuchazao') and *Arabidopsis thaliana*. Black regions indicate consensus of amino acids.

### *3.2. Protein prediction*

Using ProtParam tools, The predicted molecular weight (MW) of CsPR1 protein is 17552.59 kDa, with theoretical isoelectric point (pI) of 4.96, extinction coefficient of 37275 $M^{-1}\cdot cm^{-1}$ (assuming all pairs of Cys residues form cystines), instability index of 31.37 (considered as stable), aliphatic index of 71.94 and grand average of hydropathicity of -0.229. The strongest prediction (reliability class = 1) indicates that CsPR1 is a secreted protein and contains a signal peptide through secretory pathways, where cleavage site may be located between position 24 (Ala) and 25 (Gln). The secondary structure of CsPR1 is predicted to possess 30% alpha helixes, 15.62% extended strands and 54.37% random coils. Based on the pair-wise structure similarity by I-TASSER, The three-dimensional structural model of CsPR1 was exhibited in Fig.4, revealing 5 α-helixes from N terminal to C terminal (Fig.4a) and 2 ligand binding site residues (His72 and His118, Fig.4b) in complex with $Zn^{2+}$. Alignment to match the model to all structures in the PDB library showed CAP-domain protein from the human hookworm parasite *Necator americanus* (PDB entry 3NT8) and pathogenesis-related protein P14a (PDB entry 1CFE) had the top 2 closest structural similarity [19, 20]. Similarly, it was reported that the conserved His residues are capable of directly coordinating a zinc ion as observed in CAP domain [19]. It contains four alpha-helixes and a short helix in the pathogenesis-related protein P14a, consistent to our findings [20].

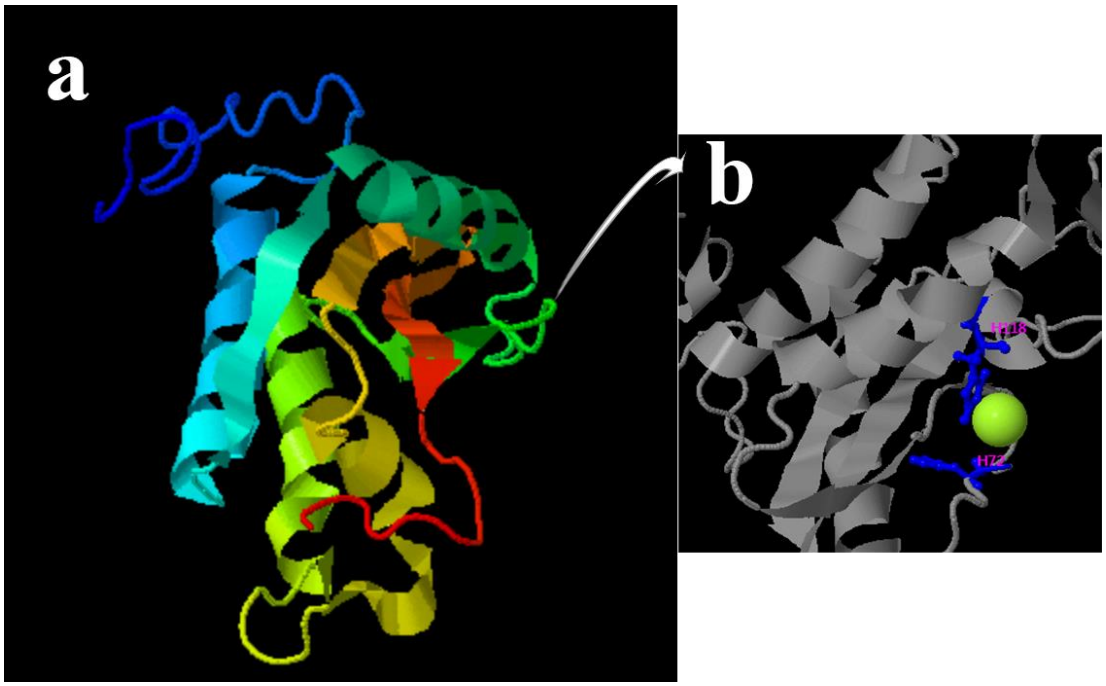



**Fig. 4.** (a) Tertiary structure of CsPR1 protein based on I-TASSER. The colors from blue to red represents N-terminal to C-terminal. (b) Zinc ion binding sites, shown in a yellow ball.

*3.3. Relative expression analysis*

To detect the expression of *CsPR1* in cultivars FD and LJ43, quantitative real-time PCR (qRT-PCR) was performed. As shown in Fig.5, relative expression level of *CsPR1* in the anthracnose infected leaves is significantly higher than healthy leaves among two cultivars. The increasement of *CsPR1* level after infection in FD is remarkably higher than that in LJ43, suggesting that *CsPR1* acts as an important guideline for resistance to pathogens. Previous study in tobacoo demonstrated that constitutive high-level expression of PR-la resulted in tolerance to infection by *Peronospora tabacina* and *Phytophthora parasitica* var. *nicotianae* [21]. The characterized PR1 proteins elevated expression and exhibited high fungicidal activity in *Arabidopsis thaliana*, tomato, cereals, ect. [22-24]. These confirmed our findings reasonable.

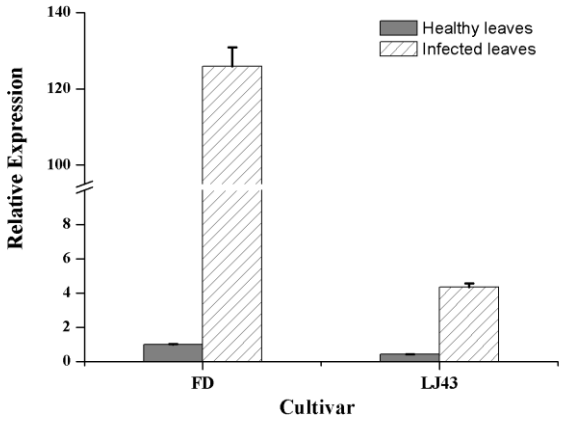



**Fig. 5.** Relative expression analysis of CsPR1 in FD and LJ43.

## 4. CONCLUSIONS

In conculsion, the cloned *CsPR1* consists of 483 bp ORF encoding a secreted protein with 160 amino acids. CsPR1 protein containing a CAP domain may act as anti-fungal agents through sequestrating sterols from pathogens to inhibit pathogen growth in the extracellular space. The upregulation of *CsPR1* occurred when tea plant was infected by anthracnose was stronger in resistant tea cultivar than that in anthracnose suspectable cultivar. It is concluded that the CsPR1 protein can be used as a defence marker and distinguishable features for identifying resistance of tea cultivars to anthracnose.

## ACKNOWLEDGMENTS

The authors appreciate the Science Technology Department of Zhejiang Province (Project No. 2016C02053-5) and the Science Technology Bureau of Ningbo City (Project No. 201701CX-D01065) for financial support to this research.

## REFERENCES

1. Linthorst, HJM 1991, "Pathogenesis-Related Proteins Of Plants", *Critical Reviews In Plant Sciences,* vol. 10, no. 2, pp. 123-150.
2. Van Loon, LC 2006, "Significance of inducible defense-related proteins in infected plants", *Annual Review Of Phytopathology,* vol. 44, pp. 135-162.
3. Fu, ZQ 2013, "Systemic Acquired Resistance: Turning Local Infection into Global Defense", *Annual Review Of Plant Biology, Vol 64,* vol. 64, pp. 839-863.

4. Selitrennikoff, CP 2001, "Antifungal proteins", *Applied And Environmental Microbiology,* vol. 67, no. 7, pp. 2883-2894.
5. Van Loon, LC 1999, "The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins", *Physiological And Molecular Plant Pathology,* vol. 55, no. 2, pp. 85-97.
6. Breen, S 2017, "Emerging Insights into the Functions of Pathogenesis-Related Protein 1", *Trends In Plant Science,* vol. 22, no. 10, pp. 871-879.
7. Uknes, S 1992, "Acquired-Resistance In Arabidopsis", *Plant Cell,* vol. 4, no. 6, pp. 645-656.
8. Liu, W 2016, "Research Progress on Anthracnose of Tea Plant. (in Chinese)", *Chinese Journal of Tropical Agriculture,* vol. 36, no. 11, pp. 20-26.
9. Cannon, PF 2012, "Colletotrichum - current status and future directions", *Studies In Mycology,* no. 73, pp. 181-213.
10. Yoshida, K 2010, "Induction of Disease Resistance in Tea (Camellia sinensis L.) by Plant Activators", *Jarq-Japan Agricultural Research Quarterly,* vol. 44, no. 4, pp. 391-398.
11. Wang, L 2016, "Transcriptome Analysis of an Anthracnose-Resistant Tea Plant Cultivar Reveals Genes Associated with Resistance to Colletotrichum camelliae", *Plos One,* vol. 11, no. 2.
12. Kumar, S 2016, "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets", *Molecular Biology And Evolution,* vol. 33, no. 7, pp. 1870-1874.
13. Yang, JY 2015, "The I-TASSER Suite: protein structure and function prediction", *Nature Methods,* vol. 12, no. 1, pp. 7-8.
14. Gao, Y 2019, "Analysis of Differentially Expressed Genes in Tissues of Camellia sinensis during Dedifferentiation and Root Redifferentiation", *Scientific Reports,* vol. 9.
15. Livak, KJ 2001, "Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method", *Methods,* vol. 25, no. 4, pp. 402-408.
16. Schneiter, R 2013, "The CAP protein superfamily: function in sterol export and fungal virulence", *Biomol Concepts,* vol. 4, no. 5, pp. 519-25.
17. Gamir, J 2017, "The sterol-binding activity of PATHOGENESIS-RELATED PROTEIN 1 reveals the mode of action of an antimicrobial protein", *Plant Journal,* vol. 89, no. 3, pp. 502-509.
18. Felsenstein, J 1985, "Confidence-Limits on Phylogenies - an Approach Using the Bootstrap", *Evolution,* vol. 39, no. 4, pp. 783-791.
19. Asojo, OA 2011, "Structure of a two-CAP-domain protein from the human hookworm parasite Necator americanus", *Acta Crystallographica Section D-Biological Crystallography,* vol. 67, pp. 455-462.
20. Fernandez, C 1997, "NMR solution structure of the pathogenesis-related protein P14a", *Journal Of Molecular Biology,* vol. 266, no. 3, pp. 576-593.
21. Alexander, D 1993, "Increased tolerance to two oomycete pathogens in transgenic tobacco expressing pathogenesis-related protein 1a", *Proc Natl Acad Sci U S A,* vol. 90, no. 15, pp. 7327-31.
22. Bowling, SA 1994, "A Mutation In Arabidopsis That Leads To Constitutive Expression Of Systemic Acquired-Resistance", *Plant Cell,* vol. 6, no. 12, pp. 1845-1857.
23. Niderman, T 1995, "Pathogenesis-Related Pr-1 Proteins Are Antifungal - Isolation And Characterization Of 3 14-Kilodalton Proteins Of Tomato And Of a Basic Pr-1 Of Tobacco With Inhibitory Activity Against Phytophthora-Infestans", *Plant Physiology,* vol. 108, no. 1, pp. 17-27.

24. Muthukrishnan, S 2001, "Pathogenesis-related proteins and their genes in cereals", *Plant Cell Tissue And Organ Culture,* vol. 64, no. 2-3, pp. 93-114.