Proceedings of the International Society of Sugar Cane Technologists, volume 29, 1747-1750, 2016

Identification of candidate secretory effector proteins (CSEPS) genes from *Colletotrichum falcatum* and their role in host-pathogen interaction by comparative modelling

CN Prasanth[1], R Viswanathan[1], P Malathi[1], AR Sundar[1] and T Tiwari[2]

[1]*ICAR-Sugarcane Breeding Institute, Coimbatore 641007, Tamil Nadu, India*; rasaviswanathan@yahoo.co.in
[2]*Xcleris Genomics Pvt. Ltd, Ahmadabad 380015, India.*

Abstract Protein-protein interaction during host-pathogen interface plays a vital role in initiating infection and is considered to be critical in understanding mechanisms of infection and defense response. *Colletotrichum falcatum,* an ascomycete fungal pathogen, infects sugarcane stalks and causes red rot. To decipher its molecular signature and pathogenicity, the genome and transcriptome of *C. falcatum* were sequenced using Illumina Hi-Seq 2000. The *C. falcatum* genome (48.2 Mb) was assembled *de novo* using the software SSpace & Velvet and the transcriptome (31 Mb) was assembled using Soap *de novo*. The genome of *C. falcatum* consisted of 12,270 genes, while its transcriptome revealed 13,320 genes, similar to the closely related species *C. graminicola* and *C. sublineola* that cause anthracnose in maize and sorghum, respectively. About 884 classical secretory proteins and 56 transmembrane helices were predicted from the *C. falcatum* transcriptome using SignalP. The *C. falcatum* secretory proteins have large number of CAZy families, esterase, peptidases, proteases, cytochrome P450, transporters, proteinase and transcription factors. An *in silico* prediction of proteins involved in host-pathogen interaction has been carried out using Cytoscape, which determined the pathways involved in establishing pathogen colonization and signal transduction, String Tool was used to predict functional protein involved in host pathogen interaction. The comparative modeling of protein-protein recognition from host-pathogen interaction and CSEPs predictions from genomic and transcriptomic context brought new understanding of the functions of effector proteins in host defense and fungal pathogenicity during the interaction.

Key words Host-pathogen interaction, effectors, CSEPs, secretory proteins

INTRODUCTION

Sugarcane is an important crop grown mostly in tropical and subtropical regions and it supplies more than 75% of the world's sugar consumption. It has the capacity to store large amounts of sucrose in the stem. Currently, it is also considered as the most efficient bioenergy crop for bio-ethanol production. One important factor impairing sugarcane productivity is the severity of some diseases affecting the crop. Among the various diseases caused by fungi, bacteria, nematodes and viruses in sugarcane, red rot, caused by *Colletotrichum falcatum*, has been a constant concern for sugar industry in India. This disease accounts for a nationwide loss of 5-10% in cane yield, making it the most serious threat among the biotic stresses. In severe epidemics, up to 100% yield losses have been recorded (Viswanathan, 2010). Currently, epidemic occurrences of red rot are noticed in popular varieties such as CoS8436, CoSe92423, CoSe95422, CoLk8102 in sub-tropical regions. In the tropical region, sugarcane production is constrained in the states of Tamil Nadu, Gujarat, Andhra Pradesh and Orissa due to red rot (Viswanathan 2013, 2016). Since red rot has been a major productivity constraint for the past 100 years, serious efforts have been made to tackle the problem by cultivation of resistant varieties. However, the varieties succumb to the newly emerging pathogenic variants of *C. falcatum*, often referred to as 'varietal breakdown', and this caused elimination of many elite varieties (Viswanathan 2016). Hence, stable resistance to red rot could not be achieved during the past decades in sugarcane.

The genome and transcriptome of *C. falcatum* were sequenced to decipher genes/mechanisms associated with pathogenesis and virulence. Although among various *Colletotrichum* fungi, the genome sizes and class of genes are similar, disease symptoms and host specificity are unique to each *Colletotrichum* species/plant interaction (O'Connell *et al.* 2012; Baroncelli *et al.* 2014; Gan *et al.* 2013). Therefore, comparative analyses of genomes may help to uncover genes that trigger specific host responses. The effector proteins from hemibiotrophic fungi interact with host proteins to carry out most of the biological functions such as signal transduction, protein transport, immune response and other pathogenic functions (Davis *et al.* 2007; Gan *et al.* 2010; Kleemann *et al.* 2012; Stergiopoulos and de Wit 2009). The strategy utilized

by most of fungal pathogens to invade into their respective host system is through secretory effectors. These secretory effectors will recognize host receptor for host-pathogen interactions.

We have analyzed the candidate secretory effector proteins (CSEPs) that are present in *C. falcatum.* This protein class was compared with those of other *Colletotrichum* spp. to determine their uniqueness in invading into the host system. Interaction between hemibiotrophs and their host plants are considerably more complex and subtle, and the CSEPs actively suppress host and allows invasive hyphae spread throughout to initiate infection. Identifying the crucial key genes that interact with host interface using *in silico* techniques will help us to understand the pathogen mechanism and mode of transmission during host-pathogen interaction.

METHODOLOGY

We used total RNA of *Cf671* fungal mat to prepare cDNA and constructed a paired-end sequencing library using Illumina TruSeq SBS Kit v3 as per the manufacture's protocol. The libraries of all samples were in the size range of 200-600 bp. The cDNA library prepared was sequenced on the Illumina HiSeq 2000 (Genotypic solutions, Bengaluru (Genome), Xcleris Lab Ltd, Ahmedabad (Transcriptome), India). Both ends of the cDNA were sequenced and the 90 bp pair-end raw reads were generated by the Illumina Genome Analyzer II system. Clean reads were obtained by removing the empty reads, the adaptor sequences, and the low-quality sequences (reads with unknown base pairs 'N'). The clean reads were then assembled into contigs and scaffolds based on pair-end information using SOAP De novo and VELVETTE. Finally, gaps of the scaffolds were filled using paired-end to obtain unigenes that contained the least number of 'N's and could not be extended on either end. The generated high quality, alignable sequence data was achieved by a 2x100 bp library with >150x coverage in genome and >100x coverage in transcriptome. The CSEPs from the *C. falcatum* transcriptome were predicted using SignalP. We have applied a computational whole-genome protocol (Fig. 1) that generates testable predictions of host-pathogen-protein interactions using Cytoscape and String. This sequence-based approach has identified several genes that are putatively involved during interaction as the first line of defense in the host and virulence-associated genes in the pathogen.

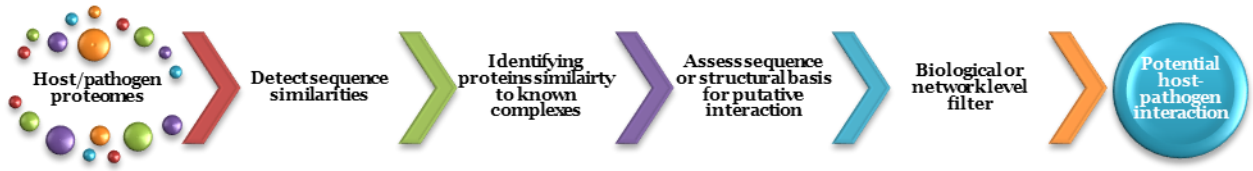

Fig. 1. Flowchart of the methodology for predicting CSEPs genes.

We analyzed 12,270 predicted genes present in the *C. falcatum* genome and 13,320 genes in the transcriptome for the presence of signal peptides using SignalP4.1 (Petersen *et al.* 2011) and localization of all signal peptides containing proteins was predicted using targetP with the default parameters. The amino acid sequences with positive SignalP prediction for signal peptide cleavage site at N-terminal region were predicted by TMHMM (Krogh *et al.* 2001) and were selected as the candidate secreted proteins. The predicted secretome was also searched for similarity across all sets of effectors using BLASTx.

RESULTS

We found 798 protein sequences which have a signal peptide among which 30 contained transmembrane segments from the genome and 884 classical secretory proteins from the transcriptome. The proteins containing signal peptides were

localized with the default parameters, which resulted in 739 sequences to the secretory pathway, 27 to the mitochondrion and 2 chloroplast signal peptides from the *C. falcatum* genome. (Fig. 2).

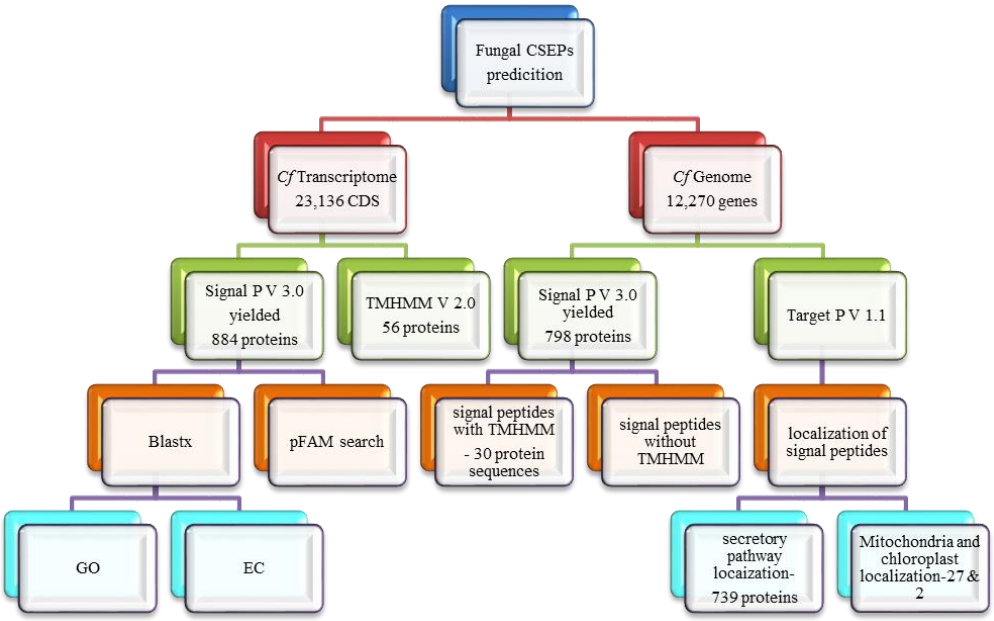

Fig. 2. Classification of CSEPs identified from genome and transcriptome of *C. falcatum*.

These secretory proteins have a large number of CAZy genes, esterase, proteinase, peptidases, cytochrome P450s, transporters and transcription factors. We classified the CSEPs into apoplastic and cytoplasmic effectors. These apoplastic and cytoplasmic effectors will be further characterized *in planta* to determine their expression during host-pathogen interface using qRT-PCR.

CONCLUSION

The repertoire of *C. falcatum* effectors predicted from the genome and transcriptome identified several putative genes that are probably involved in biotrophy-necrotrophy transition in functionally diverse patterns. This will facilitate further analysis of stage specific genes, fungal pathogenicity determinants and the expression of CSEPs *in planta*.

ACKNOWLEDGEMENTS

We thank Dr Bakshi Ram, Director of the Institute, for encouragement and facilities.

REFERENCES

Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon M. 2014. Draft genome sequence of *Colletotrichum sublineola* a destructive pathogen of cultivated sorghum. *Genetics Announcements* 2(3): 10–11.

Davis FP, Barkan DT, Eswar N, McKerrow JH, Sali A. 2007. Host pathogen protein interactions predicted by comparative modeling. *Protein Science* 16: 2585-2596.

Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell R J, Narusaka Y, Shirasu K. 2013. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytologist* 197: 1236-1249.

Gan PHP, Rafiqi M, Hardham AR, Dodds PN. 2010. Effectors of biotrophic fungal plant pathogens. *Functional Plant Biology* 37: 913-918.

Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, van Themaat EVL, O'Connell R.J. 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathogen* 8: e1002643.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* 305: 567-580.

O'Connell RJ, Thon MR, Hacquard S, *et al.* 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nature Genetics* 44: 1060-1055.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8: 785-786.

Stergiopoulos I, de Wit PJ. 2009. Fungal effector proteins. *Annual Review of Phytopathology* 47: 233-263.

Viswanathan R. 2010. *Plant Disease: Red Rot of Sugarcane.* Anmol Publishers, New Delhi. p. 306.

Viswanathan R. 2013. Sustainable ecofriendly disease management systems in sugarcane production under the changing climate. *Journal of Mycology and Plant Pathology* 43: 12-27.

Viswanathan R. 2016. Pathogen virulence in sugarcane red rot pathogen versus varieties in cultivation: classical case of loss in virulence in the pathotype CF06 (Cf671). *Sugar Tech* doi: 10.1007/s12355-016-0458-6.

Identification de gènes de CSEP (protéines candidates effectrices de sécrétion) chez *Colletotrichum falcatum* et leur rôle dans l'interaction hôte-pathogène par modélisation comparative

Résumé. L'interaction protéine-protéine durant l'interface hôte-pathogène joue un rôle vital dans l'initiation de l'infection et elle est considérée comme déterminante pour comprendre les mécanismes de l'infection et de la réponse de défense. *Colletotrichum falcatum*, champignon ascomycète pathogène, infecte les tiges de canne à sucre et cause la morve rouge. Afin de déchiffrer sa signature moléculaire et son pouvoir pathogène, le génome et le transcriptome de *C. falcatum* ont été séquencés en utilisant Illumina Hi-Seq 2000. Le génome de *C. falcatum* (48,2 Mb) a été assemblé *de novo* grâce au logiciel SSpace & Velvet et le transcriptome (31 Mb) a été assemblé *de novo* en utilisant Soap. Le génome de *C. falcatum* était constitué de 12.270 gènes, tandis que son transcriptome en révélait 13.320, ce qui est similaire chez les espèces étroitement apparentées *C. graminicola* et *C. sublineola* responsables respectivement de l'anthracose du maïs et du sorgho. Environ 884 protéines sécrétoires classiques et 56 hélices transmembranaires ont été prédites à partir du transcriptome de *C. falcatum* en utilisant SignalP. Les protéines sécrétoires de *C. falcatum* sont composées d'un grand nombre de familles de la banque de données CAZy, des estérases, des peptidases, des protéases, du cytochrome P450, des transporteurs, des protéinases et des facteurs de transcription. Une prédiction *in silico* des protéines impliquées dans l'interaction hôte-pathogène a été faite en utilisant Cytoscape, ce qui a permis de préciser les voies impliquées dans la colonisation de la plante par le pathogène et la transduction du signal. L'outil String Tool a été utilisé pour prédire la protéine fonctionnelle impliquée dans l'interaction hôte-pathogène. La modélisation comparative de la reconnaissance protéine-protéine associée à l'interaction hôte-pathogène, et les prédictions de CSEP basés sur les contextes génomique et transcriptomique ont apporté une nouvelle compréhension des fonctions des protéines effectrices dans la défense de l'hôte et dans le pouvoir pathogène du champignon au cours de l'interaction.

Mots-clés: Interaction hôte-pathogène, effecteurs, CSEP, protéines sécrétrices

La identificación de genes candidatos de proteínas secretoras efectoras (CSEPS) de *Colletotrichum falcatum* y su papel en la interacción huésped-patógeno mediante el modelado comparativo

Resumen. La interacción proteína-proteína durante la interfaz de huésped-patógeno desempeña un papel fundamental en el inicio de la infección y se considera que es fundamental para entender los mecanismos de infección y de respuesta de defensa. *Colletotrichum falcatum*, un hongo patógenico de los ascomicetos, infecta los tallos de caña de azúcar y causa la pudrición roja. Para descifrar su identificación y patogenicidad molecular, el genoma y transcriptoma de *C. falcatum* fueron secuenciados utilizando Illumina Hi-Sec 2000. El genoma de *C. falcatum* (48.2 Mb) fue ensamblado *de novo* utilizando el software SSpace & Veldet y el transcriptoma (31 Mb) se ensambló usando el método SOAP*denovo*. El genoma de *C. falcatum* consistió de 12.270 genes, mientras que su transcriptoma reveló 13.320 genes, similar al de la especie estrechamente relacionada *C. graminicola* y *C. sublineola*, que causan la antracnosis en maíz y sorgo, respectivamente. Cerca de 884 proteínas clásicas de secreción y 56 transmembranares-hélice se predijeron a partir del transcriptoma de *C. falcatum* usando SignalP. Las proteínas secretorias de *C. falcatum* tienen un gran número de familias CAZy, esterasa, peptidasa, proteasa, citocromo P450, transportadoras, proteinasa y factores de transcripción. Una predicción *in silico* de las proteínas implicadas en la interacción hospedero-patógeno se han llevado a cabo utilizando Cytoscape, que determinó las vías involucradas en el establecimiento de la colonización del patógeno y transducción de señal. Se utilizó la herramienta ¨STRING¨ para predecir las proteínas funcionales que intervienen en la interacción hospedero-patógeno. El modelado comparativo de reconocimiento proteína-proteína de la interacción huésped-patógeno y las predicciones CSEPs a partir del contexto genómico y transcriptómico, trajo una nueva comprensión de las funciones de las proteínas efectoras en la defensa del hospedero y la patogenicidad del hongo durante la interacción.

Palabras clave: Interacción planta-patógeno, efectores, CSEPs, proteinas secretorias