Sugar Tech (Nov-Dec 2017) 19(6):604–615
DOI 10.1007/s12355-017-0529-3

RESEARCH ARTICLE

Unraveling the Genetic Complexities in Gene Set of Sugarcane Red Rot Pathogen *Colletotrichum falcatum* Through Transcriptomic Approach

C. Naveen Prasanth[1] · R. Viswanathan[1] · Neethu Krishna[2] · P. Malathi[1] · A. Ramesh Sundar[1] · Tanushree Tiwari[3]

Received: 18 March 2017 / Accepted: 13 May 2017 / Published online: 29 May 2017

Abstract The transcriptome-based gene set expression of a unique sugarcane stalk infecting fungal pathogen reveals novels insights in deciphering the class of pathogenicity genes present in *Colletotrichum falcatum*. This study gains significance in understanding the genetic signature of this pathogen using RNA-Seq technology. A total of 53,410,513 reads (24,732 transcripts) specific to *C. falcatum* were generated, and 13,320 genes were predicted. Gene ontology distributions have been grouped into three domains as biological (3053), cellular (1601) and molecular functions (3444). KEGG annotations represented pathway biomolecules such as carbohydrates, lipids, nucleotides, amino acids, glycans, cofactors, vitamins, terpenoids and polyketides. The genes for virulence have been classified and grouped into candidate effectors, transition-specific and secondary metabolites, proteases, transporters and peptidases which revealed that *C. falcatum* transcripts encode a large number of secondary metabolites and membrane transporters. Gene enrichment analysis revealed that the number of transporters encoded by *C. falcatum* is significantly more as compared to that encoded by several other *Colletotrichum* spp. Phylogenomics analysis indicated that *C. falcatum* is closely related to *C. graminicola* and *C. sublineola* infecting related host plants, maize and sorghum, respectively. This study provides a comprehensive understanding of *C. falcatum* pathobiology and has identified many candidate genes/putative functions possibly required for its pathogenesis .

Keywords *Colletotrichum falcatum* · Sugarcane · RNA-Seq analysis

Introduction

Sugarcane (*Saccharum* spp. hybrid) is an economically important crop cultivated in the tropical and subtropical regions across the world and contributes to nearly 70% of global white sugar requirements (http://www.statista.com/statistics/249679/total-production-of-sugar-worldwide, accessed November 12, 2014; http://www.sucden.com/statistics/7_white-sugar-trade, accessed November 12, 2014). This also emerges as an alternate source for bioenergy since ethanol from sugarcane is used as automobile biofuel in Brazil and other countries. In addition, bagasse a by-product of sugar industry is used to generate electricity and utilized as a raw material for paper production. In India, sugarcane is cultivated in around 4.32 M ha and the country ranks second after Brazil in terms of cultivable area and sugar production (http://www.vsisugar.com/india/statistics/world_indiasugar.html, accessed November 12, 2014). In India, ~580 sugar mills depend on sugarcane as the raw material to produce sugar, ethanol and electricity. The crop contributes to 2% of country's GDP and 8–10% of country's agricultural GDP (http://www.vsisugar.com/india/statistics/world_indiasugar.html, accessed November

Electronic supplementary material The online version of this article (doi:10.1007/s12355-017-0529-3) contains supplementary material, which is available to authorized users.

✉ R. Viswanathan
rasaviswanathan@yahoo.co.in

[1] Division of Crop Protection, Sugarcane Breeding Institute, Indian Council of Agricultural Research, Coimbatore 641007, India

[2] Indian Institute of Spices Research, Indian Council of Agricultural Research, Calicut, Kerala 673012, India

[3] Xcleris Genomics Pvt. Ltd., Ahmadabad, Gujarat 380009, India

12, 2014). Under Indian conditions, the crop production is severely affected by different diseases and in particular by red rot, caused by the fungus *Colletotrichum falcatum* Went (Ascomycotina, Perfect stage: *Glomerella tucumanensis* [Speg.] Arx & Muller) (Viswanathan 2010). The occurrence of severe red rot epidemics in sugarcane is attributed to the frequent emergence of new virulent pathotypes of the fungus (Viswanathan and Rao 2011). In India, new sugarcane varieties are released for cultivation based on their yield potential, sugar content and essentially with red rot resistance. Replacement of established cane varieties due to its susceptibility to red rot affects the sustainability of the farmers and the sugar industry. During the last 100 years, several disease epidemics have been reported in the country and many high sugar varieties such as Co 419, Co 997, Co 1148, Co 6304, Co 7805, CoC 671, CoC 85061, CoC 92061, CoJ 64 and CoS 8436 were withdrawn from cultivation owing to red rot, thus leading to great economic losses in sugar industry. Sugarcane is a complex polypoid, and inheritance of red rot resistance is poorly understood (Viswanathan 2010). Moreover, *C. falcatum* pathotypes of Indian origin possess/exhibit genetic variation, and the pathogenic variation could be mainly due to gradual adaptation of the existing pathotypes to the newly introduced varieties in the field (Malathi et al. 2006, 2010). Unlike other *Colletotrichum* hemibiotrophs, *C. falcatum* infects stalk tissues in sugarcane, which ultimately leads to drying of foliage tissues, but occasionally, the pathogen infects leaves, causing only lesions on midrib and infects laminar tissues. The pathogen infects the stalk and converts the stored sucrose into glucose and fructose through inversion, which makes the infected canes deprived for sugar extraction (Viswanathan 2010). The stalk infection causes lesions in leaves and subsequently drying of whole stalk, and in severe cases, the entire field dries up (Fig. 1a). Internally, the entire parenchymatous tissues turn reddish with horizontal white patches (Fig. 1b); pith regions show longitudinal cavities very often filled with dark greyish mycelium of the fungal pathogen (Viswanathan and Samiyappan 1999). The intriguing fungal pathogen has made several epidemics in India, and limited information is available to study the establishment of fungal pathogenesis. Except for a few attempts on studying the pathogen variability at molecular level in *C. falcatum* (Suman et al. 2005; Malathi et al. 2010), not much work has been done in India to understand the molecular basis of virulence or pathogenicity mechanisms. Further, complete characterization of the pathogenic fungus with genomic information is very much necessary, and there is a need to generate transcriptome data of the fungus to understand its pathogenicity mechanism.

New types of sequencing technology have brought a revolutionary insight in large genome/transcriptome sequencing which is capable of generating large data in identifying biological signature of an organism (Mardis 2013). This recent advancement made in next-generation sequencing (NGS) platforms precisely RNA-Seq technology has brought in revolutionary insights in deciphering gene arrangements, expression patterns and functional interaction among gene families in many important plant pathogens. NGS has made a landmark achievement in sequencing draft genome/transcriptome using short reads and exploited multiple-level sequencing (Baroncelli et al. 2014; Gan et al. 2013; O'Connell et al. 2012; Wu et al. 2013). The RNA-Seq technology generated remarkable outcome in sequencing several organisms, and de novo assembly of short reads becomes very promising and generated ample information which helps to study those organisms completely (Morozova and Marra 2008; Lister et al. 2009; Wang et al. 2009; Wilhelm and Landry 2009). Here we described the *C. falcatum* transcriptome and obtained genomic insights on this pathogen using Illumina Hi-Seq2000. A total of 24,732 contigs were assembled from *C. falcatum* transcriptome; 13,320 genes were predicted and putatively characterized their functions. This study also predicted the secretome, effectors and genes putatively involved in virulence. Also the phylogenomic study of *Colletotrichum* spp. with respect to other ascomycete fungal pathogens has been carried out. This is the first detailed report on transcriptome of *C. falcatum* using the NGS platform, the results of which suggest that *C. falcatum* has a close synteny with *C. graminicola* and *C. sublineola* whose lifestyles have been deciphered (O'Connell et al. 2012; Baroncelli et al. 2014). Unraveling the genes involved in *C. falcatum* has given a comprehensive understanding of its pathogen biology and their putative genes functions. The present study has opened up new vistas in making a beginning to look into the pathogenicity mechanism of *C. falcatum*.

Materials and Methods

Fungal Culture Used

The *C. falcatum* pathotype *Cf*671 (isolated from the sugarcane variety CoC 671) (MTCC accession number-12142) maintained at the red rot culture collection facility of the institute was used for the current study. The fungus was multiplied on oatmeal agar (oatmeal 30 g, agar 20 g, water l L) for 7 days. Monoconidial culture of the fungus was inoculated on complete media broth (CMB) and incubated

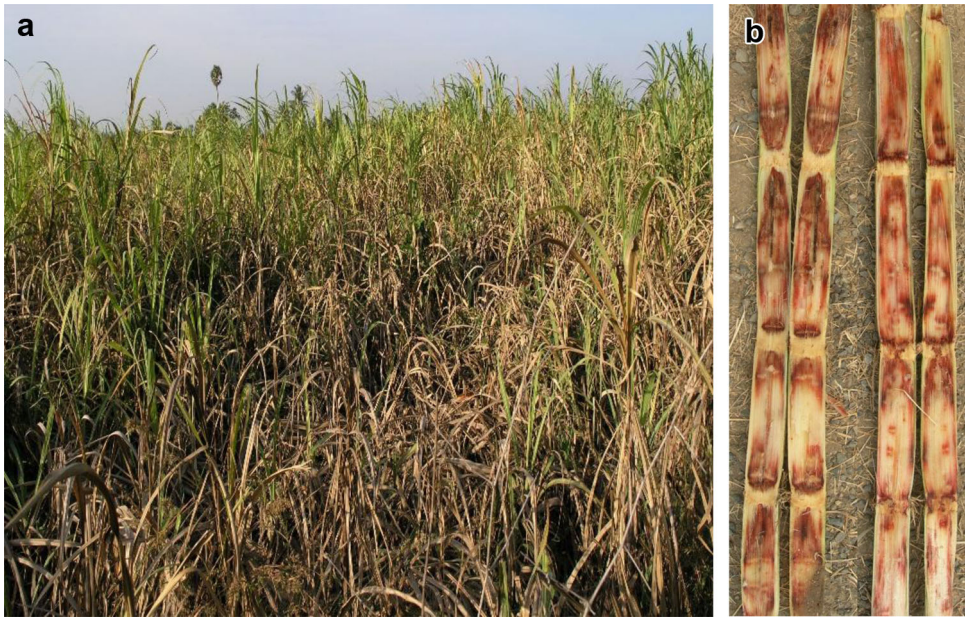

Fig. 1 Damage caused by ascomycete *C. falcatum* in sugarcane. **a** Sugarcane crop exhibiting severe red rot infection in the field (cv Co 92012, Nagapattinam District, Tamil Nadu State, India) and **b** internally infected stalks showing tissue damage and discoloration

under room temperature (28 ± 2 °C) for 12 days (without shaking) for RNA isolation.

RNA Extraction and Library Preparation for Transcriptome Sequencing

Total RNA was extracted from *C. falcatum Cf*671 using TRI Reagent (Sigma-Aldrich, USA) and treated with Rnase-free DNAse I (Promega, USA). Subsequently, the quality of RNA was checked in 1% denatured agarose gel electrophoresis for the presence of intact 28 and 18S bands, and RNA was quantified using Nanodrop-8000. The paired-end cDNA sequencing library was prepared using Illumina TruSeq SBS Kit v3 as per the manufacture's protocol. Library preparation was started with mRNA fragmentation followed by reverse transcription, second-strand synthesis, paired-end adapter ligation and finally ended with index PCR amplification of adaptor-ligated library. Library quantification and quality check were performed on Caliper Lab Chip GX using HT DNA High Sensitivity Assay Kit. The libraries of all samples were in the size range of 200–600 bp. The resulting libraries were validated using the Agilent Bio Analyzer 2100 onto the Agilent High Sensitivity Chip performed at Xcleris Genomics Pvt. Ltd, Ahmadabad, India.

Cluster Generation and Sequencing Run

Paired-end sequencing allows the template fragments to be sequenced in both forward and reverse directions. Cluster generation was carried out by the hybridization of template molecules onto the oligonucleotide-coated surface of the flow cell. Immobilized template copies were amplified by bridge amplification to generate clonal clusters. This process of cluster generation was performed on cBot using TruSeq PE Cluster kit v3-cBot-HS. The kit reagents were used for binding of samples to complementary adapter oligos on paired-end flow cells. The adapters were designed to allow selective cleavage of the forward cDNA strand after resynthesis of the reverse strand during sequencing. The copied reverse strand was then used to sequence from the opposite end of the fragment. TruSeq SBS v3-HS kits were used to sequence cDNA of each cluster on a flow cell using sequencing by synthesis technology on the HiSeq 2000 done at Xcleris Genomics Pvt. Ltd. Ahmadabad, India.

Transcriptome Sequencing, Assembly and Annotation

The cDNA library prepared was sequenced on the Illumina HiSeq 2000 performed by Xcleris Genomics Pvt. Ltd. Ahmadabad, India, and both ends of the cDNA were sequenced. Clean reads were obtained by removing the empty reads, the adaptor sequences and the low-quality sequences (reads with unknown base airs 'N') using Trimmomatic V 0.33. The clean reads were then assembled into contigs and transcripts based on paired-end information using CLC genomics workbench V 6.0 on default parameters. The CDS regions were predicted using Trans Decoder V 2.0.1. The *C. falcatum* transcriptome was annotated using BLASTX and BLAST2GO. The KEGG pathway annotation was performed using KAAS automated server.

Gene Enrichment Analysis and Protein Family Classification

GO sequence distribution helped in specifying all the annotated nodes comprising of GO functional groups. CDS of transcripts associated with similar functions were assigned to the same GO functional group. The GO sequence distributions were analyzed for all the three GO domains, i.e., biological processes, molecular functions and cellular components. Protein families were classified by searching the assembled transcripts against Pfam (Punta et al. 2012) and InterProScan (Quevillon et al. 2005). Protease families were identified using BLASTp against MEROPS peptidase database release 9.6 (Rawlings et al. 2012). Cytochromes (CYPs) were named according to classification details collected from BLASTp against fungal cytochrome P450 database version 1.2 (Park et al. 2008). Carbohydrate-degrading enzymes selected from InterProScan and Pfam analyses were classified according to GH (Glycoside hydrolase) family as classified in the CAZy database (Cantarel et al. 2009). Proteases, carbohydrate-degrading enzymes and membrane transporters were predicted from *C. falcatum* transcriptome and compared with *C. graminicola, C. sublineola C. higginsianum, C. gloeosporioides* and *C. orbiculare* (UniProt database available: http://www.uniprot.org/downloads, accessed August 14, 2014).

KEGG Pathway Identification

Orthologous assignment and mapping of the transcripts to the biological pathways were performed using KEGG automatic annotation server (KAAS) (Moriya et al. 2007). All the transcript contigs were compared against the KEGG database (Kanehisa and Goto 2000; Kanehisa et al. 2012) using BLASTx with threshold bit-score value of 60. The mapped transcript contigs represented metabolic pathways of major biomolecules such as carbohydrates, lipids, nucleotides, amino acids, glycans, cofactors, vitamins, terpenoids and polyketides. The mapped contigs also represented the genes involved in genetic and environmental information processing and cellular processes.

Secretome Analyses

The putative candidate effectors prediction is a challenging task with multiple strategies involved in finding both classical and non-classical proteins. The amino acid sequences with positive SignalP (Petersen et al. 2011) 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 putatively predicted genes were also searched for similarity across all set of effectors using BLASTx.

SSR (Simple Sequence Repeats) Identification

SSR mining or microsatellites were identified using MISA databases (http://pgrc.ipk-gatersleben.de/misa/, accessed September 4, 2014) with certain number of criteria to avoid interruption of compound microsatellite localization. *C. falcatum* transcripts were analyzed to mine potential SSR, which were defined as di- to hexa-nucleotide SSR with a minimum of four repeats for all motifs (Annot8r program available: http://www.nematodes.org/bioinformatics/annot8r/index.shtml, accessed August 30, 2014).

Results

Statistical Analysis and Validation of *C. falcatum* Transcriptome Sequences

A total number of 56,637,987 raw reads were generated, and from that, 53,410,513 high-quality reads were filtered out and assembled using CLC genomics workbench on default parameters, and a total 24,732 distinct contigs were obtained. We finally obtained transcripts of 31.19 MB size ranging from 200 to 1000 bp (Supplementary Fig. 1). The size distribution of these contigs is shown in Table 1.

Functional Annotation and Classification of *C. falcatum* Transcriptome

The functional annotation was performed using the 23,136 predicted CDS from transcript contigs of *C. falcatum* by aligning the CDS and annotated using BLASTX and BLAST2GO. The *C. falcatum* CDS resulted in annotation of 19,454 from 23,136 CDS, while 3682 CDS had no significant BLAST hits. The species distribution of the best match result for each sequence revealed that 9046 genes have best matches (first hit) with *C. graminicola*, followed by *C. sublineola* (8125), *C. higginsianum* (1158), *C. gloeosporioides* (465), *C. orbiculare* (295), *Nectria haematococca* (161), *Verticillium albo-atrum* (77) *Magnaporthe oryzae* (41), *Chaetomium globosum* (41), *Metarhizium acridum* (28), *Neurospora crassa* (16) and *Ustilago maydis* (1) (Fig. 2).

Gene-Enriched Pathway and Classification

A total of 23,136 predicted CDS, which were assigned with at least one GO term, were categorized into 23 functional groups (Fig. 3). The genes involved in cellular process representation of genes involved in intracellular,

Table 1 Assembly statistics of *C. falcatum* transcripts

Description	Cf-tR
Number of transcript contigs	24,732
Sum transcript contigs length (in base pairs)	31,194,966
Max transcript contigs length (in base pairs)	24,916
Min transcript contigs length (in base pairs)	200
N50 value (in base pairs)	2337

extracellular, macromolecules and membrane-bound complex were found dominant. The transcription regulator, binding, catalytic activity and enzyme regulator represented high implication of genes involved in molecular process. The largest group was 'multi-organismal process,' followed by 'pigmentation' and 'signal transduction mechanisms,' and the smallest groups were 'metallochaperone' and 'extracellular structures' that represented the development of organism 'biological process' (Supplementary Fig. 2).

KEGG Pathway Mapping

A total of 2399 unigenes were predicted as pathway regulatory genes and were annotated in the KAAS database, assigned to several KEGG pathways (Fig. 4). The 'metabolic pathway' comprised of a large proportion containing 1289 unigenes occupying 53.73%, along with 'biosynthesis of secondary metabolites,' 'xenobiotics biodegradation and metabolism.' The 'genetic and cellular pathway' contained 969 unigenes occupying 45.39%, along with 'folding, sorting and degradation' and environmental adaptation.

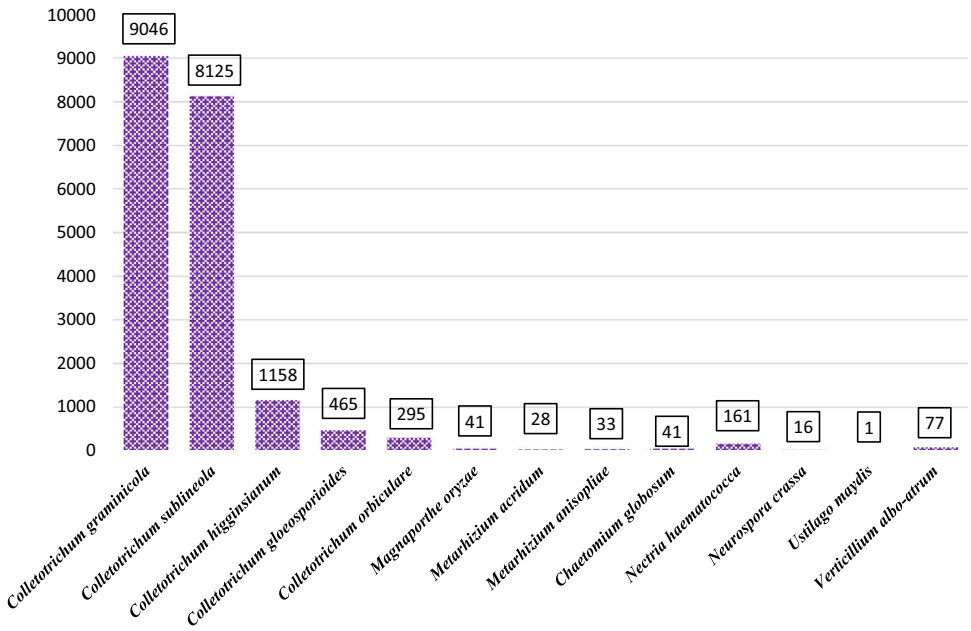

Fig. 2 Phylogenomic analysis and gene synteny between *C. falcatum* transcriptome with other *Colletotrichum*

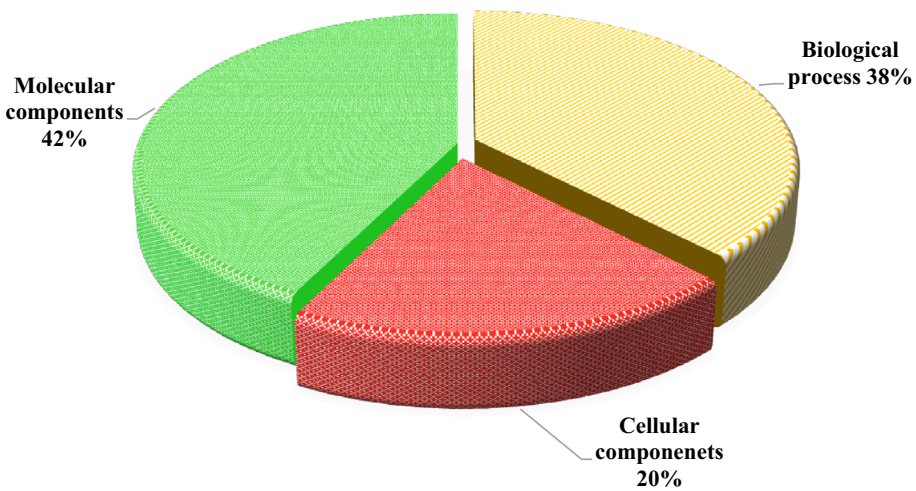

Fig. 3 Gene ontology (GO) distribution—from the predicted CDS of *C. falcatum* unigenes, classified into three major components (molecular components, biological process and cellular components)

Protein Family Classification

The protein families were classified based on Pfam and InterProScan search, and 16 classes of protein families have been predicted for fungal pathogenesis in *C. falcatum* (Fig. 5). The size of various protein families has been compared with ascomycota and other classes of fungal transcriptomes including *Aspergillus nidulans, Botrytis cinerea, Cordyceps militaris, Fusarium graminearum, Magnaporthe oryzae, Metarhizium anisopliae, Metarhizium acridum, Neuropora crassa, Sclerotinia sclerotiorum* and *Ustilaginoidea virens*. The 'major facilitator superfamily,' 'lipase' and 'cytochrome P450' class of protein families were found abundant, and esterase/thioesterase group of protein families were found to be present in *C. falcatum* in large numbers (Fig. 5). The glycoside

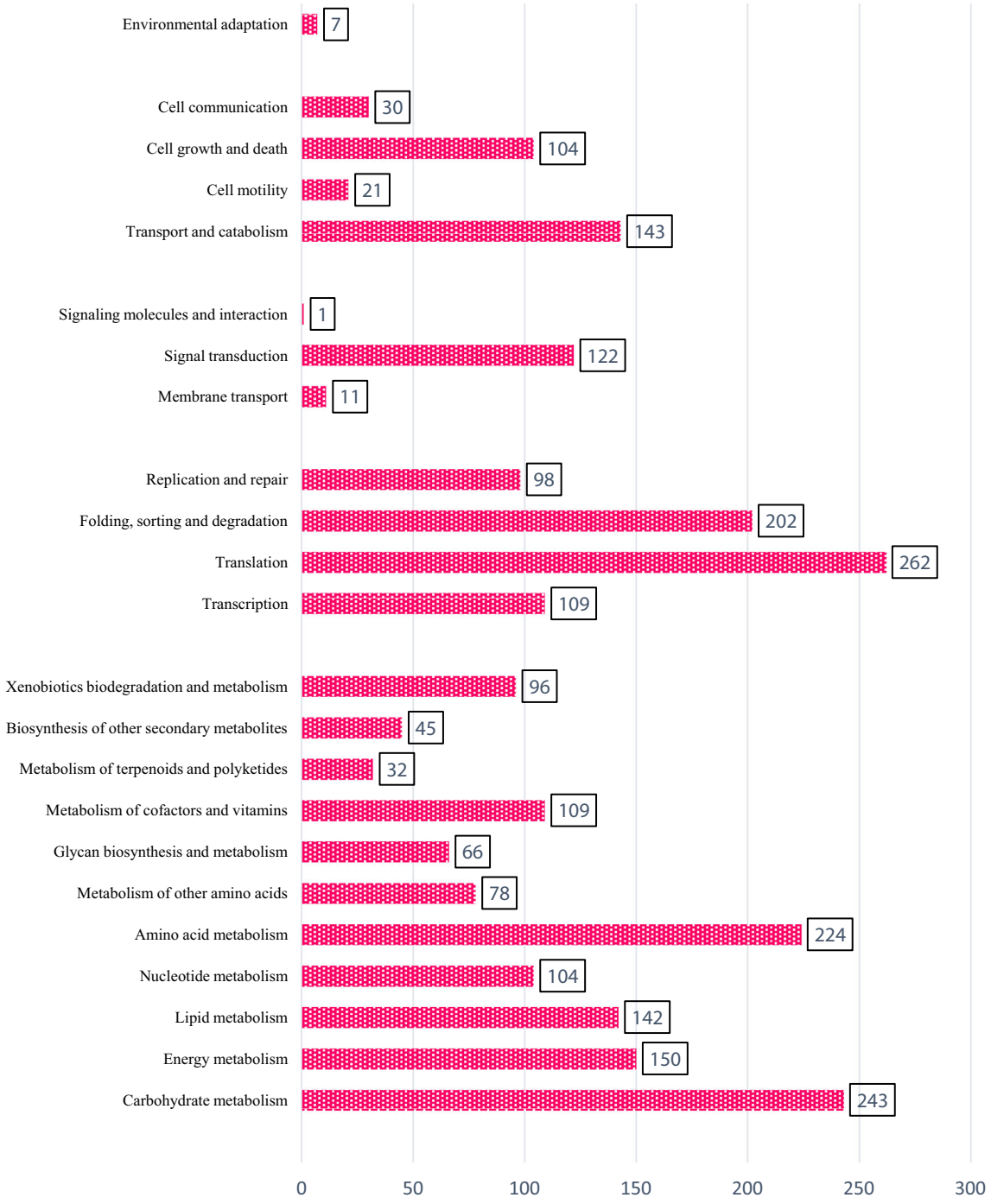

Fig. 4 *Colletotrichum falcatum* transcriptome annotations and distributions of expressed genes in KEGG pathway

hydrolase (GH) genes of *C. falcatum* (203) are quiet close to other ascomycete fungi (Table 2).

Peptidase Family Classification

The peptidase families were predicted from MEROPS peptidase database, and this led to identification of 10 major categories of peptidases in *C. falcatum* (Fig. 6). Glutamic peptidase, serine peptidase, metallopeptidases and threonine peptidases constituted the majority of peptidase families present in *C. falcatum*. A total number of 219 genes were predicted as peptidases in *C. falcatum,* as well in other *Colletotrichum* spp. such as *C. sublineola*, *C. orbiculare, C. gloeosporioides* and *C. graminicola* (Supplementary Fig. 3).

Identification of Predicted Secreted Proteins and Its Enzymatic Function

In *C. falcatum*, about 884 classical secretory proteins were found to have signal peptides and those peptides were defined as secretome of *C. falcatum*. The secretory proteins were analyzed for transmembrane domains to define the presence of transmembrane helices using TMHMM. About 56 transmembrane helices were present in predicted signal peptides, which are further referred to as secreted proteins with transmembrane region. The secreted proteins and 56 transmembrane helices were classified using GO, which revealed that signal peptides have a major role in stabilizing fungal secretory proteins in the host system during pathogenesis. These secretory proteins were found to have large number of CAZy families, esterase, peptidases, proteases, cytochrome P450, transporters, proteinase and transcription factors which will be further validated at *in planta* level.

Membrane Transporter Classification

The *Colletotrichum* group of fungi expresses large number of transporters, and those transporters were classified based on BLASTp search against Transporter Classification Database (www.tcdb.org, accessed September 6, 2014). About 625 membrane transporters were found in *C. falcatum*, and the classes of transporters were compared with other hemibiotrophic and necrotrophic pathogens (Table 3).

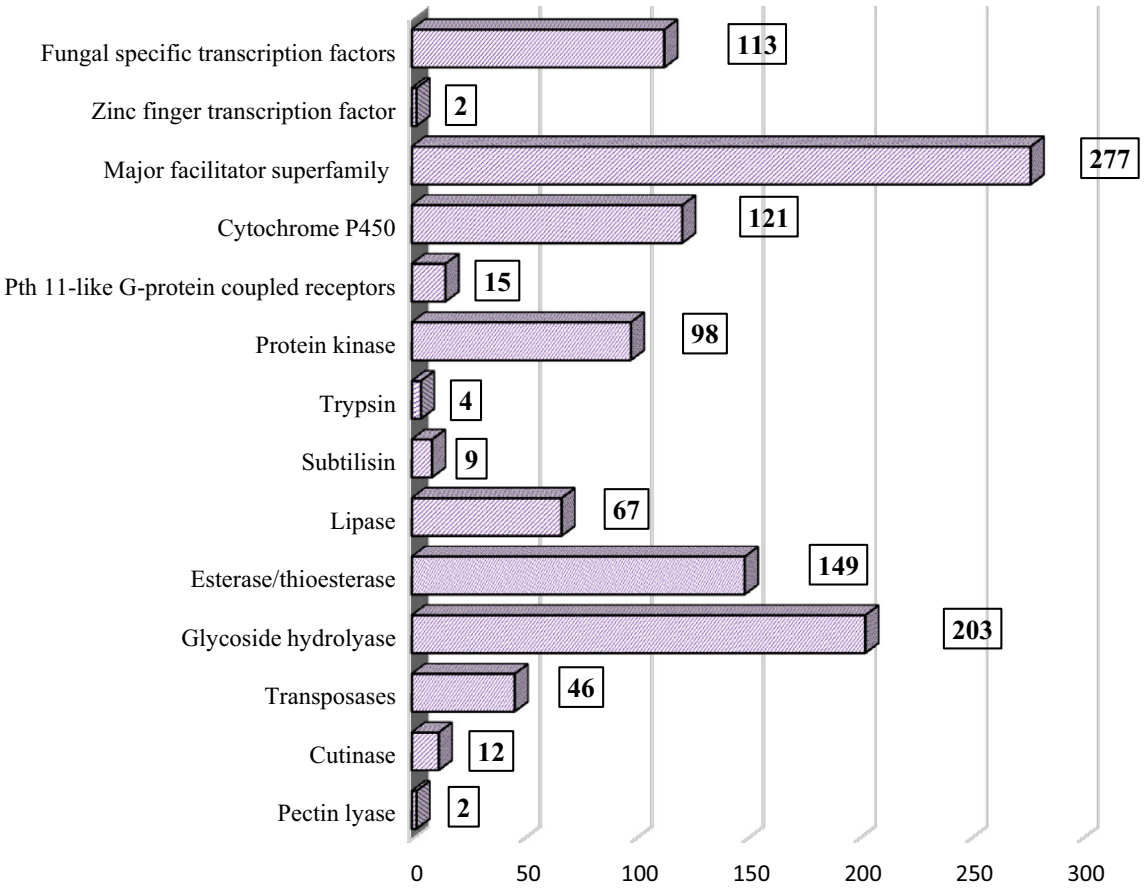

Fig. 5 Protein families involved in fungal pathogenesis in *Colletotrichum falcatum*

Table 2 Selected protein families involved in fungal pathogenesis in *C. falcatum* and other ascomycete fungus

Protein family	CF	UV	FG	MO	BC	SS	CCM	MAA	MAC	NC	AN
Fungal-specific transcription factors	113	116	253	95	114	95	157	165	161	88	258
Zinc-finger transcription factors	2	25	85	58	48	54	63	86	53	63	58
Major facilitator superfamily	277	80	246	180	190	144	207	215	177	106	261
Cytochrome P450	121	38	112	137	129	93	57	123	95	40	116
Pth11-like G-protein-coupled receptor	15	17	51	50	23	25	21	29	22	20	44
Protein kinases	98	149	149	135	127	169	167	161	192	119	139
Trypsin	4	4	3	3	1	1	12	32	17	2	2
Subtilisin	9	11	24	29	7	6	35	55	43	10	4
Aspartic protease	8	14	18	19	14	21	21	33	25	19	16
Lipase	67	14	31	23	28	25	38	47	41	16	27
Pectin esterase	0	2	3	1	5	5	0	0	0	1	3
Glycoside hydrolase	203	117	204	228	158	162	126	151	123	168	231
Glycosyl transferase	56	66	91	102	93	79	82	109	94	76	90
Dehydrogenase	191	176	439	315	342	289	288	351	309	235	458
Cutinase	12	4	12	18	11	8	4	2	2	3	4
Pectin lyase	2	0	25	9	25	20	10	7	7	5	24

CF, *Colletotrichum falcatum*; AN, *Aspergillus nidulans*; BC, *Botrytis cinerea*; CCM, *Cordyceps militaris*; FG, *Fusarium graminearum*; MO, *Magnaporthe oryzae*; MAA, *Metarhizium anisopliae*; MAC, *Metarhizium acridum;* NC, *Neuropora crassa*; SS, *Sclerotinia sclerotiorum*; and UV, *Ustilaginoidea virens*

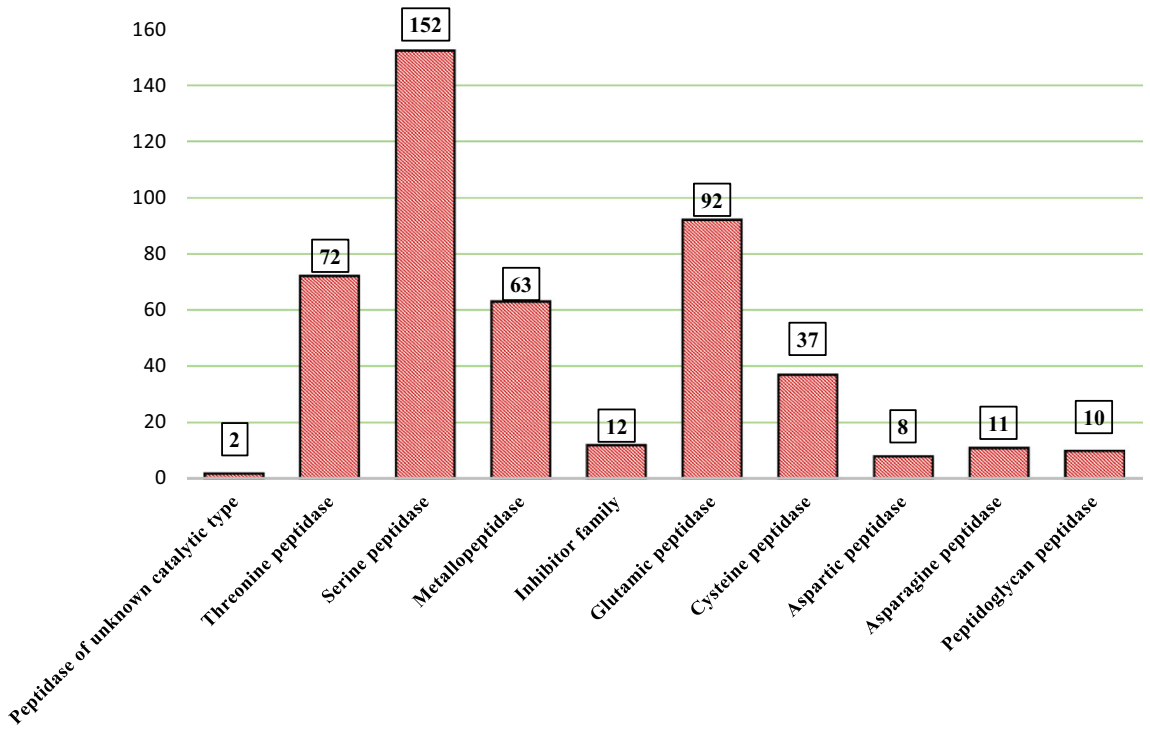

Fig. 6 Peptidase families involved in fungal pathogenesis in *Colletotrichum falcatum*

Secondary Metabolite Predictions

The secondary metabolites are the key precursors in contributing virulence to fungal pathogens in the host system. We found abundance of four major classes of secondary metabolites in *C. falcatum* transcriptome. The transcripts of secondary metabolites have been putatively classified into the following classes including polyketide synthase (PKS),

Table 3 Major transporter families involved in fungal pathogenesis in *C. falcatum* and other ascomycete fungus

Family abbreviation	CF	CG	CH	SC	MG	BC	VA	BG	UM	FO	LB	MP	MA
ATP-binding cassette superfamily	41	52	68	23	42	49	47	10	30	75	37	23	91
Pleiotropic drug resistance family	11	14	20	12	10	17	18	4	10	20	11	14	22
Broad specificity multidrug resistance efflux pump	4	9	9	4	9	8	12	3	8	9	9	4	9
Major facilitator superfamily	277	299	363	77	224	221	248	46	83	447	82	71	422
Sugar porter family	58	66	94	29	60	58	83	10	18	134	14	19	79
Drug:H + Antiporter-1 (12 Spanner) family	59	64	61	13	51	64	40	14	24	100	26	16	120
Drug:H + Antiporter-2 (12 spanner) family	32	37	43	9	39	44	23	10	11	52	9	8	52
Siderophore–iron transporter family	2	4	10	6	3	5	10	0	3	13	1	2	20
Monocarboxylate porter Family	10	19	25	4	19	23	25	0	8	38	9	3	52
Anion:Cation Symporter	74	82	104	10	52	33	67	5	14	115	13	23	81
Oligopeptide transporter family	20	16	18	3	12	12	8	4	7	14	10	20	30
Proton-dependent oligopeptide transporter family	1	5	7	1	4	3	2	1	1	6	2	1	8
Amino acid–polyamine–organocation family	36	32	34	19	28	29	29	13	15	62	22	12	56
Total	625	699	856	210	553	566	612	120	232	1085	245	216	1042

CF, *Colletotrichum falcatum*; CG, *C. graminicola*; CH *C. higginsianum*; SC, *Saccharomyces cerevisiae*; MG *Magnaporthe grisea*; BC, *Botrytis cinerea*; VA, *Verticillium albo-atrum*; BG, *Blumeria graminis* f. sp. *hordei*; UM, *Ustilago maydis*; FO, *Fusarium oxysporum*; LB, *Laccaria bicolor*; MP, *Melampsora laricipopulina*; MA, *Metahrizium anisopliae*

non-polyketide synthase (NPKS), dimethylallyl tryptophan synthase (DMATs) and terpene synthase (TS). The PKS class of genes contained three domains (iterative Type I, Type II, Type III), which contribute mainly on melanin synthesis and appressorium-mediated host penetration.

SSR Distribution and Classification

A total of 7807 simple sequence repeats (SSR) were identified from expressed transcripts, and the SSRs were distributed based on localized sizes (Table 4). The SSRs have a large number of motifs with varying types (7807), followed by dinucleotide (3583; 45.8%), trinucleotide (3652; 46.89%), tetra-nucleotide (432; 5.5%), penta-nucleotide (82; 1.5%) and hexa-nucleotide (58; 0.8%) motifs. Further analysis of the length of SSRs revealed that is mainly distributed from 6 to 10 bp which accounted for 86.7% of the total simple sequence repeats (SSR).

Discussion

The *Colletotrichum* spp. are economically important plant pathogens, causing diseases on a wide range of crops and significantly affecting crop yields (Damm et al. 2010; Cannon et al. 2012; McDowell 2013). Among the *Colletotrichum* spp., *C. falcatum* is a stalk infecting pathogen and causes severe epidemics in sugarcane. The study is to find the nature of gene set and enriched pathway of *C. falcatum* in establishing this as a stalk infecting pathogen and to decipher the unique lifestyle of this pathogen. Earlier, the genome of *C. falcatum* has been reported with 48.2 Mb size and 12,270 genes were present (Viswanathan et al. 2016) which is merely equal to the transcriptome of *C. falcatum* which has 13,320. In the present study, we have sequenced an Indian isolate of *C. falcatum—Cf*671 which has caused severe outbreaks in the field on popular varieties in the tropical India (Viswanathan 2010). We report de novo assembly and gene-enriched pathway analyses of *C. falcatum* and its in silico analyses to explore expressed gene insights to understand its pathogenicity mechanisms. The high-throughput massively parallel sequencing helped us to generate the transcriptome consensus of ~31 Mb. The successful RNA-Seq technology brought new insights on deciphering the *C. falcatum* transcriptome information with a comprehensive understanding on GO, KEGG, protein family classification, secretome identification and secondary metabolites.

The NGS technology has been applied to sequence many pathogenic fungi of *Colletotrichum* spp. such as *C. orbiculare, C. gloeosporioides, C. sublineola, C. higginsianum* and *C. graminicola* (O'Connell et al. 2012; Gan et al. 2013; Kleemann et al. 2012; Baroncelli et al. 2014). The first ascomycete genome-wide transcriptional profiling has been done in *Magnaporthe oryzae* to uncover the morphogenesis by Oh et al. (2000). Several other hemibiotrophic fungal pathogens like *Phytophthora capsici, Moniliophthora roreri, Venturia inaequalis* and *Ustilaginoidea virens* were sequenced using NGS platforms (Jupe et al. 2013; Thakur et al. 2013; Meinhardt et al. 2014;

Table 4 SSR Table 2: Short sequence repeats (SSR) statistics of *C. falcatum* transcripts

Description	Cf-tR
Total number of transcript contigs examined	24,732
Total size of transcript contigs (bp)	31,194,966
Total number of identified SSRs	7807
Number of SSR containing transcript contigs	5058
Number of transcript contigs containing more than one SSR	1650
Number of SSRs present in compound formation	717

Zhang et al. 2014). Earlier, transcriptome profiling of sugarcane smut pathogen *Sporisorium scitamineum* was done through Solexa Sequencing platforms (Wu et al. 2013). The complete genome sequence of *Sporisorium scitamineum*, biotrophic pathogen, was sequenced using Pac Bio technology and Illumina HiSeq 2000 and recently reported with transcriptome analysis of biotrophic interaction with sugarcane (Taniguti et al. 2015). The *C. falcatum* transcriptome profiling brought a new opportunity to identify the genes coding for development, protein families, secondary metabolites and transportation. Among the *Colletotrichum* spp., *C. falcatum*, *C. graminicola* and *C. sublineola* are the specialized pathogens to infect stalk tissues in the monocots and have ability to protrude the host environment completely, whereas other *Colletotrichum* spp. infect leaves, twigs, fruits, etc. (Stephenson et al. 2005; Casado-Díaz et al. 2006; Münch et al. 2008; Gan et al. 2013; O'Connell et al. 2012).

Colletotrichum falcatum transcriptome revealed expression of pathogenicity-associated genes belonging to four major classes (candidate effectors, transporters, CAZy and secondary metabolites) which has brought a revelation in elucidating pathogenicity mechanisms of *C. falcatum*. These pathogenicity determinants have specifically been involved in successful invasion of the host plant (O'Connell et al. 2012). The *C. falcatum* has a large number of organic compounds involved in signal transduction or degradation, namely secondary metabolites, MFS transporters, fungal-specific transcription factors, protein kinases and proteases families (Figs. 5, 6, Supplementary Fig. 3). Further, the expanded classes of genes, viz. cytochrome P450, lipases, glycoside hydrolases and polyketide synthase families for detoxification, hydrolysis or secondary metabolites biosynthesis are significantly larger in number than *C. graminicola* (O'Connell et al. 2012). The RNA-Seq of *C. orbiculare* revealed the presence of expanded class of genes, which has been a revelation of hemibiotrophic lifestyle in *Colletotrichum* (Gan et al. 2013), whereas the presence of class of genes and proteins in *C. falcatum, C. sublineola, C. gloeosporioides, C. higginsianum* and *C. graminicola* is found to be analogous. The fungal colonization and development of morphogenesis happen inside the host through transportation of superfamily complexes which can be referred as effectors (O'Connell et al. 2012). The candidate secretory effector proteins (CSEPs) of *C. falcatum* have a large number of crucial effectors, which mediate host–pathogen interaction, and those signal peptides could immobilize plant hormones, which may lead to suppression of host resistance against pathogen infection (Houterman et al. 2008; Stergiopoulos and de Wit 2009). Gene set enrichment analysis revealed a large set of proteinases and peptidases which possibly determine the pathogenicity of *C. falcatum* (Figs. 5, 6). In addition, these expressed genes may be involved in the production of fungal toxins and melanin, which would contribute toward virulence and successful invasion of the fungal pathogen into the host system. During the interactions of fungal pathogens with their host plants, pathogens undergo several complex and crucial phases, including attachment to the plant surface, germination on the plant surface and formation of infective structures, penetration into the host and colonization of the host tissue (Schafer 1994; Stergiopoulos and de Wit 2009). The release of certain pathogenicity factors like toxin, melanin, cell wall-degrading enzymes (pectinase, cellulase, hemicellulase, protease, amylase and phospholipase), cutinase and hormones are essential in pathogenesis process. Interestingly, many of these proteins are identified in *C. falcatum* in the present study, and most of these protein families were involved in the hemibiotrophic stage shift (O'Connell et al. 2012, Gan et al. 2013) (Fig. 5) (Table 4). Furthermore, many gene-encoded proteins such as G-coupled protein receptors, kinases, core genes for the biosynthesis of secondary metabolites, CYPs, ATP-binding cassette (ABC) transporters and MFS transporters are directly or indirectly involved in these processes.

There are about 625 and 699 membrane transporter genes in *C. falcatum* and *C. graminicola*, respectively (Table 4). Over all, it suggests that the *C. falcatum* membrane transportation system is close to *C. graminicola.* The ABC transporters and the MFS transporters play a vital role in establishing pathogenicity, and these two big superfamilies enhance other transporters like sugar porters, monocarboxylate porters, anion: cation importer involved in KEGG pathways and nutrient uptake in CAZy, etc. (McDowell 2013). The ABC and MFS drug transporters can function as endogenous fungal pathogenic factors such as toxins in plant pathogenic fungi and also in protecting against exogenous plant defense compounds such as phytoalexins, thereby suppressing host immunity (Zheng et al. 2011). The presence of similar number of transporters in *C. falcatum* with other hemibiotrophic and expansion in multi-facilitator superfamily (MFS) suggests that *C. falcatum* may also adopt hemibiotrophic lifestyle like other

species of *Colletotrichum*. We found that the *C. falcatum* transcriptome encodes a large quantity of transporters and they have been compared with other sequenced ascomycete fungi (Table 4).

Fungal secondary metabolites are the key precursors in pathogenesis; fungal developmental stage represents conidial development, germ tube development and appressorial formation. These secondary metabolites present in *C. falcatum* have been classified into several classes of genes coding for secondary metabolites, namely PKS, NRPS, PKS-NRPS hybrid, DMATS and TS genes (Supplementary Fig. 4). The Type I PKS contained three domains mainly ketoacyl synthase (KS), acyl transferase (AT) and phosphopantetheine attachment site (PP or ACP). Large numbers of CoA esters have been found to be abundant in *C. falcatum* transcriptome. The Type II PKS contains domains like ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains. The Type III PKS contains chalcone synthase and flavone synthase, which were found to be an interactive isomer for facilitating host–pathogen interaction. The NPKS, DMAT and TS contributed toward production of various lethal compounds like sesquiterpene cyclases (STC), diterpene cyclases (DTC), cyclize farnesyl pyrophosphate (FPP) and geranyl pyrophosphate (GGPP) (Supplementary Fig. 4). The gene set analyses revealed that apart from CSEPs, membrane transporters and CAZy, *C. falcatum* has a potential to produce large number of secondary metabolites which are probably involved in pathogenicity. *C. graminicola* has three types of PKS genes which are considered as key precursors in pathogenesis (O'Connell et al. 2012), and these three types are found to be identical with *C. falcatum* secondary metabolites (Supplementary Fig. 4).

Conclusion

Our results suggested that gene families and other active enzymes are similar to other hemibiotrophic pathogens. The study has identified several putative pathogenicity determinants, candidate effectors, transition-specific metabolites, proteases and housekeeping genes. A large number of genes present in *C. falcatum* were found to be the genes encoding biotrophy–necrotrophy transition further that will be identified through *in planta*. The membrane transporters were identified, and comparative analyses revealed that the number of transporters encoded by *C. falcatum* is significantly more as compared to that encoded by several other *Colletotrichum* spp. infecting different crops (Gan et al. 2013; O'Connell et al. 2012). The transcriptome-based gene annotation gave a clear view that *C. falcatum* is more close to *C. graminicola* and *C. sublineola* (the causal organism of anthracnose in maize and sorghum, respectively) in sharing the similar gene set families and clusters of enriched groups, respectively. Further, gene set enrichment analyses for validation and identification of novel genes are under progress to identify virulence genes/effectors. In conclusion, the findings from this study provide a better understanding of the biology of the fungal pathogen and have identified several unique putative genes/functions required for its pathogenesis in sugarcane. Also this is the first attempt to sequence *C. falcatum* in NGS platform. The transcriptome-based gene annotation lays a foundation to comprehensively understand pathogenicity of *C. falcatum* in sugarcane.

Acknowledgements The authors are grateful to the Director of ICAR-Sugarcane Breeding Institute, Coimbatore, for providing facilities and encouragement.

Funding This study was partly funded by Outreach Project of ICAR, ALCOCERA.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Standards This article does not contain any studies with human participants or animals performed by any of the authors.

Data Access All data contributing to this transcriptome initiative have been deposited at the NCBI under BioProject PRJNA272832. The accession number of Sequence Read Achieves (SRA) is SRR1765657.

References

Baroncelli, R., J.M. Sanz-martín, G.E. Rech, S.A. Sukno, and M. Thon. 2014. Draft genome sequence of *Colletotrichum sublineola* a destructive pathogen of cultivated sorghum. *Genome Announcement* 2 (3): 10–11.

Cannon, P.F., U. Damm, P.R. Johnston, and B.S. Weir. 2012. *Colletotrichum*—Current status and future directions. *Studies in Mycology* 73: 181–213.

Cantarel, B.L., P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research* 37: D233–D238.

Casado-Díaz, A., S. Encinas-Villarejo, B.D.L. Santos, E. Schilirò, E.M. Yubero Serrano, F. Amil Ruíz, and J.L. Caballero. 2006. Analysis of strawberry genes differentially expressed in response to *Colletotrichum* infection. *Physiologia Plantarum* 128: 633–650.

Damm, U., R. Barroncelli, L. Cai, Y. Kubo, R.J. O'Connell, B. Weir, K. Yoshino, and P.F. Cannon. 2010. *Colletotrichum*: Species, ecology and interactions. *IMA Fungus* 1 (2): 161–165.

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

Houterman, P.M., B.J. Cornelissen, and M. Rep. 2008. Suppression of plant resistance gene-based immunity by a fungal effector. *PLoS Pathogens* 4: e1000061.

Jupe, J., R. Stam, A.J. Howden, J.A. Morris, R. Zhang, P.E. Hedley, and E. Huitema. 2013. *Phytophthora capsici*-tomato interaction features dramatic shifts in gene expression associated with a hemi-biotrophic lifestyle. *Genome Biology* 14: R63.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28: 27–30.

Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. 2012. KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Research* 40: D109–D114.

Kleemann, J., L.J. Rincon-Rivera, H. Takahara, U. Neumann, E.V.L. van Themaat, and R.J. O'Connell. 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathogens* 8 (4): e1002643.

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

Lister, R., B.D. Gregory, and J.R. Ecker. 2009. Next is now: New technologies for sequencing of genomes, transcriptomes, and beyond. *Current Opinion in Plant Biology* 12 (2): 107–118.

Malathi, P., R. Viswanathan, and R. Jothi. 2006. Specific adaptation of *Colletotrichum falcatum* pathotypes to sugarcane cultivars. *Sugar Tech* 8: 52–56.

Malathi, P., R. Viswanathan, A. Ramesh Sundar, N. Prakasam, P. Padmanaban, R. Jothi, S.R. Renuka Devi, and M. Poongothai. 2010. Variability among *Colletotrichum falcatum* pathotypes used for screening red rot resistance in sugarcane. *Sugar Cane International* 28 (2): 47–52.

Mardis, E.R. 2013. Next-generation sequencing platforms. *Annual Reviews of Analytical Chemistry* 6 (1): 287–303.

McDowell, J.M. 2013. Genomic and transcriptomic insights into lifestyle transitions of a hemi-biotrophic fungal pathogen. *New Phytologist* 197: 1032–1034.

Meinhardt, L.W., G. Gilson, L. Costa, D.P.T. Thomazella, P. José, P.L. Teixeira, and B.A. Bailey. 2014. Genome and secretome analysis of the hemibiotrophic fungal pathogen *Moniliophthora roreri*, which causes frosty pod rot disease of cacao: Mechanisms of the biotrophic and necrotrophic phases. *BMC Genomics* 15: 164. doi:10.1186/1471-2164-15-164.

Moriya, Y., M. Itoh, S. Okuda, A.C. Yoshizawa, and M. Kanehisa. 2007. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35 (Suppl 2): 182–185.

Morozova, O., and M. Marra. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255–264.

Münch, S., U. Lingner, D.S. Floss, N. Ludwig, N. Sauer, and H.B. Deising. 2008. The hemibiotrophic lifestyle of *Colletotrichum* species. *Journal of Plant Physiology* 165: 41–51.

O'Connell, R.J., M.R. Thon, S. Hacquard, S.G. Amyotte, J. Kleemann, M.F. Torres, and L.J. Vaillancourt. 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nature Genetics* 44: 1055–1060.

Oh, Y., N. Donofrio, H. Pan, S. Coughlan, D. Brown, S. Meng, and R.A. Dean. 2000. Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus *Magnaporthe oryzae*. *Genome Biology*. doi:10.1186/gb-2008-9-5-r85.

Park, J., S. Lee, J. Choi, K. Ahn, B. Park, S. Kang, and Y.H. Lee. 2008. Fungal cytochrome P450 database. *BMC Genomics* 9: 402.

Petersen, T.N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods* 8: 785–786.

Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clement, A. Heger, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman, and R.D. Finn. 2012. The Pfam protein families database. *Nucleic Acids Research* 40: D290–D301.

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez. 2005. InterProScan: Protein domains identifier. *Nucleic Acids Research* 33: 6W116–6W120.

Rawlings, N.D., A.J. Barrett, and A. Bateman. 2012. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* 40: D343–D350.

Schafer, W. 1994. Molecular mechanisms of fungal pathogenicity to plants. *Annual Review of Phytopathology* 32: 461–477.

Stephenson, S., C.M. Stephens, D.J. Maclean, and J.M. Manners. 2005. CgDN24: A gene involved in hyphal development in the fungal phytopathogen Colletotrichum gloeosporioides. *Microbiological Research* 160: 389–397.

Stergiopoulos, I., and P.J. de Wit. 2009. Fungal effector proteins. *Annual Review of Phytopathology* 47: 233–263.

Suman, A., S. Lal, A.K. Shasany, A. Gaur, and P. Singh. 2005. Molecular assessment of diversity among pathotypes of *Colletotrichum falcatum* prevalent in sub-tropical Indian sugarcane. *World Journal of Microbiology and Biotechnology* 21: 1135–1140.

Taniguti, L.M., P.D.C. Schaker, J. Benevenuto, L.P. Peters, G. Carvalho, A. Palhares, and C.B. Monteiro-Vitorello. 2015. Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with sugarcane. *PLoS ONE* 10 (6): e0129318. doi:10.1371/journal.pone.0129318.

Thakur, K., V. Chawla, S. Bhatti, M.K. Swarnkar, J. Kaur, R. Shankar, and G. Jha. 2013. De novo transcriptome sequencing and analysis for *Venturia inaequalis*, the devastating apple scab pathogen. *PLoS ONE*. doi:10.1371/journal.pone.0053937.

Viswanathan, R. 2010. *Plant disease: Red rot of sugarcane*, 306. New Delhi: Anmol Publishers.

Viswanathan, R., and G.P. Rao. 2011. Disease scenario and management of major sugarcane diseases in India. *Sugar Tech* 13: 336–353.

Viswanathan, R., and R. Samiyappan. 1999. Red rot disease in sugarcane: A major constraint for the Indian sugar industry. *Sugar Cane* 17: 9–15.

Viswanathan, R., C.N. Prasanth, P. Malathi, and A.R. Sundar. 2016. Draft genome sequence of *Colletotrichum falcatum*—A prelude on screening of red rot pathogen in sugarcane. *Journal of Genomics* 4 (10): 1–3. doi:10.7150/jgen.13585.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10 (1): 57–63.

Wilhelm, B.T., and J.R. Landry. 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Nature Methods* 48: 249–257.

Wu, Q., L. Xu, J. Guo, Y. Su, and Y. Que. 2013. Transcriptome profile analysis of sugarcane responses to *Sporisorium scitamineum* infection using Solexa sequencing technology. *BioMed Research International* Article ID 298920.

Zheng, P., Y. Xia, G. Xiao, C. Xiong, X. Hu, S. Zhang, H. Zheng, Y. Huang, Y. Zhou, S. Wang, G.P. Zhao, X. Liu, R.J. St Leger, and C. Wang. 2011. Genome sequence of the insect pathogenic fungus Cordyceps militaris, a valued traditional Chinese medicine. Genome Biology 12(11):R116. doi: 10.1186/gb-2011-12-11-r116.

Zhang, Y., K. Zhang, A. Fang, Y. Han, J. Yang, M. Xue, and W. Sun. 2014. Specific adaptation of *Ustilaginoidea virens* in occupying host florets revealed by comparative and functional genomics. *Nature Communications* 5: 3849. doi:10.1038/ncomms4849.