

# Transcriptome Analysis of Sugarcane in Response to Colletotrichum Falcatum Infection Reveals Repertoire of Transcription Factors and Host Targets

#### C. Naveen Prasanth

ICAR Sugarcane Breeding Institute

Viswanathan Rasappa (**I** rasaviswanathan@yahoo.co.in)

ICAR Sugarcane Breeding Institute https://orcid.org/0000-0002-7274-8144

P. Malathi

ICAR Sugarcane Breeding Institute

A.R. Sundar

ICAR Sugarcane Breeding Institute

#### Research Article

**Keywords:** Transcriptome, Sugarcane, red rot, Next Generation Sequencing, Transcription factors, Colletotrichum falcatum

Posted Date: November 22nd, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-1075135/v1

**License:** © ① This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

#### **Abstract**

**Background** - Sugarcane (*Saccharum* spp hybrid), an important C4 perennial plantation crop, globally grown for white sugar and ethanol production. Red rot caused by *Colletotrichum falcatum* is one of the most important threats affecting sugarcane productivity in many countries including India.

**Materials and Methods** - Comprehensive understanding is very much needed to define their transcription level differences and their key regulatory genes during interaction of sugarcane with *C. falcatum*. To compute and evaluate the molecular mechanism in sugarcane, transcriptome analysis of sugarcane challenged with *C. falcatum* was sequenced using Hi-Seq 2500 and gene expression profiles were generated by qRT-PCR assays in both compatible and incompatible interactions after challenge inoculation of *C. falcatum* in sugarcane.

**Results** - A total of 15,728,914 reads were aligned to 48,935 unigenes using BOWTIE 2; the unigenes were annotated using BLASTX and found that 39,895 unigenes were annotated and 22,025 were unigenes with respect to host species, 8,830 with respect to *Colletotrichum* spp and 9,040 were found to be novel genes. A total of 243 transcription factors (TFs) were found to be predicted in sugarcane challenged with *C. falcatum* and those TFs were divided into 45 specific families. WRKY, MYB, NAC, bHLH and AUX/IAA transcription factors were found to be abundant which are considered to be key regulators in controlling wide range of molecular events such as defense response, oxidative stimuli, host signalling and triggering disease resistance. In addition, a lot of stress related genes and genes involved in gene ontological and KEGG pathway were significantly affected due to *C. falcatum* infection. Quantative real time PCR assays carried out to validate reliability of observed expression patterns in sugarcane in response to *C. falcatum* infection illustrates first transcriptome wide *in planta* identification and analysis of TF repertoire in the host pathogen interaction.

**Conclusion** - The results of this study provide a benchmark discovery in finding host targets and provide tissue specific data set of genes that express in response to *C. falcatum* in sugarcane and also a complete analysis of main group of genes that significantly enriched under this condition. This is the first comprehensive work provides basis for the further studies to dissect role of TFs at molecular level in sugarcane defense to fungal pathogens.

#### Introduction

Sugarcane (*Saccharum* spp) belongs to *Poaceae* family; an important cash crop cultivated worldwide for sugar and ethanol. Biotic stresses especially by plant pathogens such as bacteria, fungi and viruses are the major threat for sugarcane production across the globe. Among the pathogens, red rot of sugarcane is considered to be a serious threat and it is caused by an ascomycete pathogen *Colletotrichum falcatum* Went (Teleomorph: *Glomerella tucumanensis* [Speg.] Arx & Muller). The emergence of red rot of sugarcane has challenged sugarcane production and cultivation in India resulting considerable economic losses (1). The fungus is responsible for many severe disease outbreaks is one of the most destructive pathogens of sugarcane in India. Outbreak of red rot in sugarcane is most destructive on this annual tall grass due to

spread of the pathogen through sugarcane setts that are used as seed material (2). Several studies were conducted on the pathogen flora during the epidemics that identified emerging isolates of *C. falcatum* which is responsible break down of resistance leads to varietal failure (3,4). Unlike the most of the *Colletotrichum* spp that infect foliar tissues, twig and fruits in the plants, *C. falcatum* infects stalks tissue in sugarcane and degrade hard nodal tissue which is made of cellulose, pectin and ligninolytic substrates. Several attempts were made to understand host pathogen interaction and initially induction of pathogenesis related proteins in response to *C. falcatum* interaction (5, 6).

To understand molecular basis of the host-pathogen interaction differential display RT-PCR was initially used to identify transcripts differentially regulated during interaction in sugarcane and suspension cells (7, 8) Further, subtractive libraries were established to identify time specific and initial defense responses of sugarcane during *C. falcatum* pathogenesis (9). Besides to understand the pathogen biology in detail, the genome of *C. falcatum* was sequenced to be 48.16 MB in size with 12,270 genes (10). The transcriptome of C. falcatum (in vitro) was reported to be 31 MB with 23,136 CDS were predicted (11). Recently we have identified the candidate secretory effector proteins (CSEPs) from genome and transcriptome of C. falcatum and virulent strain specific SSR marker from the genome of C. falcatum (12, 13). The proteomicsbased core proteome analysis classification was performed during sugarcane challenged with *C. falcatum* (14). Molecular basis of breakdown of resistance and emergence of virulent pathotypes remains unexplored due to complexities in screening them in large scale. To elucidate the mechanism underlying host pathogen interaction, it is important to determine the host targets during sugarcane interaction with *C*. falcatum using next generation sequencing (NGS) technology. Sugarcane is a polyploidy crop which is complex to understand moreover the genome of sugarcane is not completely sequenced even after advent of high through put sequencing technology. With this background, transcriptome sequencing technique facilitates to elucidate host pathogen interactions especially in case of sugarcane and *C. falcatum* interaction.

In this study, we utilised NGS platform to sequence both host pathogen sequence through Illumina Hi-Seq 2500 to identify host targets in sugarcane during interaction with *C. falcatum*. For the first of its kind, a cross platform analysis was performed to find out the host targets in sugarcane during *C. falcatum* interaction using plant transcriptional factor database and found that 243 transcription factors were induced during this interaction. These results were compared with our previous findings where we attempted to exploit genes in traditional sequencing platforms.

#### **Materials And Methods**

# Plant material and pathogen inoculation

A sugarcane cultivar resistant (Co 93009) and susceptible (CoC 671) to red rot were raised by following standard cultural practices for a tropical sugarcane. The pathogenic *C. falcatum* pathotype *Cf*671 was challenge inoculated in both the cultivars as reported earlier. Sterile milliQ water was used in place of conidial suspension in mock samples. *C. falcatum* inoculated sugarcane stalks along with un-inoculated

healthy control and mock inoculated control of both the cultivars were harvested in triplicates (three independent biological samples) frozen in liquid nitrogen and stored at -80°C till RNA extraction.

# RNA extraction and library preparation for transcriptome sequencing

Total RNA was extracted from CoC 671 challenged with *Cf*671 using TRI Reagent (Sigma-Aldrich, USA) in different time intervals 2,4,6 dpi, 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 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 was 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 bp to 600bp. The resulting libraries were validated using the Agilent Bio Analyzer 2100 onto the Agilent High Sensitivity Chip performed at Nucleome Informatics Pvt. Ltd, Hyderabad, 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 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 2500 done at Nucleome Informatics Pvt. Ltd, Hyderabad, India.

# Transcriptome sequencing, assembly and annotation

The cDNA library prepared was sequenced on the Illumina HiSeq 2500 performed by Nucleome Informatics Pvt. Ltd, Hyderabad, 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 0.35. The clean reads were then assembled into contigs and transcripts based on pair-end information using Trinity and CD-HIT with default parameters. TGICL performed on 161,701 CD-HIT transcripts for analysis of large expressed sequence tags (EST) and mRNA

databases in which the sequences are first clustered based on pairwise sequence similarity to produce longer and complete consensus sequences which yielded 116,422 transcripts. Further Evidential Gene used in 116,422 transcripts to retain the biological significant transcripts.

# 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 analysed 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).

# **KEGG** pathway identification

Orthologous assignment and mapping of the transcripts to the biological pathways were performed using KEGG automatic annotation server (KAAS) (15). All the transcript contigs were compared against the KEGG database (16, 17) 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, polyketides, etc. The mapped contigs also represented the genes involved in genetic and environmental information processing and cellular processes.

# **Distribution of Transcription factors**

The transcripts were annotated against all plant transcriptional factor database protein sequences (http://planttfdb.cbi.pku.edu.cn/) to determine the transcription factors in transcriptome samples.

# Detection of simple sequence repeats (SSR)

Mining of SSR or microsatellites were predicted using MlcroSAtellite identification tool (http://pgrc.ipk-gatersleben.de/misa/) with certain number of criteria to avoid interruption of compound microsatellite localization. Sugarcane and *C. falcatum* transcripts were analysed to mine potential SSR, Compound microsatellites as well as SSR motifs are classified into Mono, di, tri, tetra, penta and hexa nucleotides.

# qRT-PCR validation for the key genes

Based on the transcriptome data, the transcription factors expressed in sugarcane challenged with *C. falcatum* WRKY, MYB, NAC and bZIP were chosen to quantitative real time polymerase chain reaction (qRT-

PCR) with a Step-One plus Real-Time PCR system (Applied Bio systems, USA) Each reaction was carried out in triplicates, the primers used for qRT-PCR were designed based on the transcriptome data except for actin and are listed in Table S3. For qRT-PCR experiments, cDNA concentration was standardized for each sample and dissociation curve analysis was performed to check primer specificity. The cocktail for qRT-PCR reaction contained 50 ng cDNA, 2.5 pmol primers, 10  $\mu$ l SYBRGREEN Master Mix and with nuclease free water the total reaction mixture was made up to 20  $\mu$ l. The reaction was performed for 40 cycles (denaturation for 10 min at 95 °C followed by annealing and extension at 1 min for 58 °C).

#### Results

# Transcriptome sequencing and assembly

The 15,728,914 clean reads of sugarcane sample assembled using Trinity transcriptome assembler with default parameters resulted to 168,960 transcripts. On the assembled 168,960 transcripts, CD-HIT-EST module from the CD-HIT was run to remove the similar short sequences based on 100% alignment coverage to longer sequence resulted in 161,701 transcripts. TGICL run using default parameters resulted to 116,422 transcripts. Evidential Gene used in 116, 422 transcripts to retain the biologically significant transcripts. Evidential Gene runs with default parameters reported 48,935 unigenes (Table 1). The sequence E-Value distribution (Fig S1) and top species from annotated unigenes were evenly distributed (Fig 1).

# Annotations and species distribution

The transcripts structural annotation performed using TransDecoder tool. TransDecoder identifies candidate coding regions within transcript sequences. It also provides information about each CDS sequences whether it is complete or partial. About 48,935 unigenes were submitted to TransDecoder, and it detected 33,661 (68.78%) unigenes with ORFs, with 8622 presenting complete ORFs. The transcriptome of sugarcane and *C. falcatum* identified 48,935 unigenes, 39,895 were found to be annotated using BlastX and among the 39,895 genes, 8,540 unigenes belonged to the fungus while 9,040 unigenes had no significant blast hits. The species distribution from the transcriptome of sugarcane and *C. falcatum* had 21,169 genes matching with *Sorghum bicolor* followed by *Zea mays* (5,000), *C. graminicola* (3250), *C. sublineola* (2980), *Setaria italica* (2400), *C. incanum* (1960), *Saccharum* cultivar R570 (1100), *C. tofieldiae* (980) and *C. higginisianum* (940) (Fig 1).

# GO analysis and distribution

The GO sequence distribution for sugarcane in response to *C. falcatum* infection was analysed for all the three GO domains i.e., biological processes, molecular functions and cellular components (Fig 2). The transcriptome of sugarcane inoculated with *C. falcatum* had 16,576 unigenes involved in biological process in which 12,445 unigenes belonged to sugarcane (Fig S2) and 4,921 unigenes with *C.* 

*falcatum* (Fig S3). The molecular functions of 17,428 unigenes in which 13,531 and 3,777 unigenes were related to sugarcane and *C. falcatum*, respectively. The cellular components contained 13,745 unigenes, in which 11,638 and 3,685 related to sugarcane and *C. falcatum*, respectively.

# **KEGG Pathway analysis and distribution**

Sugarcane and *C. falcatum* unigenes were mapped and screened for the biological pathway predictions using KAAS. The mapped unigenes represented metabolic pathways of major biomolecules such as carbohydrates, lipids, nucleotides, amino acids, glycans, cofactors, vitamins, terpenoids, polyketides, etc. The mapped unigenes also represented the genes involved in genetic information processing, environmental information processing, organismal systems and cellular processes (Fig 3).

# Distribution of transcription factors

A total of 243 TFs were found to be predicted in sugarcane challenged with *C. falcatum* and those TFs were divided into 45 specific families. WRKY, MYB, NAC, bHLH and AUX/IAA transcription factors were found to be abundant and are considered to be key regulators in controlling wide range of molecular events such as defense response, oxidative stimuli, host signalling and triggering disease resistance (Fig 4). Among the TFs, WRKY, MYB and NAC were found to present in abundance and also up regulated during interaction of sugarcane and *C. falcatum* when compared with other hemibiotrophs and necrotrophs (Fig 5).

#### Putative characterization of WRKY, NAC and MYB

WRKY transcription factor known for its regulatory role in signalling the physiological pathway involved in development or disease resistance and also plays a wide role in regulation of both biotic and abiotic stresses, the transcriptome of sugarcane challenged with *C. falcatum* showed a triggering response in up regulation of WRKY 1, 2, 11, 12, 23, 26, 33, 38, 41, 46, 48, 51, 53, 64, 69, 70 and 75. NAC gene family considered to be largest and most important plant transcription factor, NAC1, 4, 8, 11, 18, 21, 22, 23 and 67 were up regulated during sugarcane interaction with *C. falcatum*. MYB transcription factor known for its decisive role in up regulation of glucose and this TF lays foundation for increase in glucose release and abiotic stress tolerance. Several MYB TFs were found to be present in sugarcane transcriptome such as MYB 1, 4, 7, 11, 15, 21, 35, 38, 43, 46 and 108. This WRKY, NAC and MYB genes control wide range of molecular events which are putatively identified in various hemibiotrophs and necrotrophic fungal organisms during interaction with respective host systems (Table 5).

# Stress related genes

There are several stress related transcription factors expressed during sugarcane and C. falcatum interaction, AP2/ERF transcription factor involved in functional divergence during drought, salt and also for plant hormonal treatment. Production of reactive oxygen species (ROS) signalling is known earlier events in response to microbe intervention. This stress related gene was found to be expressed during sugarcane and *C. falcatum* interaction which is no wonder because ROS signalling gene triggers mostly during biotrophic necrotrophic switch (BNS) where it acts as antioxidative enzyme. Along with ROS, there are several other genes found to be expressed such as secondary metabolite, defense JA-ET signalling genes, SA signalling and Ca 2+ signalling and chalcone synthase. Pathogenesis-related (PR) proteins are pathogen or stress induced plant proteins PR 1 a primary protein which mediates SA defense response was expressed, PR 2 family (endo-1-3-β-D-glucanase) and PR 5 family (thaumatin – like) associated with pathway regulations, PR 6 (proteinase inhibitor) and PR 9 (Peroxidases) closely associated with JA-ET pathway were also found to be expressed during the interaction. Several stress related genes were manually annotated based on the results generated through various platforms like SAR priming efficacy, DD RT-PCR (Table S1), 2D based defense proteome analysis (Table S2) suggested that only a smaller number of TFs were recorded compared to transcriptome-based screening. Comparative screening revealed that large numbers of Zn2Cys6 TFs were expressed during this interaction (Table 3).

#### **Distribution of SSR or Microsatellites**

A total of 48,935 sequences were examined and 8,451 SSRs identified as unigenes with 7,425 SSR containing sequences (Table 2). There are 538 SSRs mono and tri nucleotide SSR were found to be abundant in transcripts related to sugarcane compared to *C. falcatum* (Table 2). The SSRs have a large number of motifs with varying types, Mono nucleotide (2,501; 29.5%) followed by di-nucleotide (1,131; 13.3%), tri-nucleotide (4,056; 47.9%), tetra-nucleotide (124; 1.4%), penta-nucleotide (162; 1.9%) and hexanucleotide (168; 1.9%) motifs (Fig 6) SSR motifs found to be higher in *C. falcatum* than sugarcane transcripts (Fig 6 & 7).

# Functional characterization of transcription factor during compatible and in-compatible interactions in sugarcane

The biotrophy and nectrophy switch in pathogen evokes a differential response in both resistant and susceptible cultivars of sugarcane. *C. falcatum* interaction with sugarcane triggers several signalling pathways at different phases. This transcriptional screening using qRT-PCR revealed that *C. falcatum* switches its phases easily in the susceptible cultivars but strategically switches in resistant cultivars (Fig 8). Most of the TFs were up regulated during compatible interactions but WRKY, MYB 2, bHLH 2 were down regulated during *C. falcatum* interaction with susceptible cultivar (cv CoC 671). Same way the incompatible interactions shows down regulation of multiple TFs such as WRKY 26, MYB, bZIP, bZIP2, bHLH-2 may be due to non-responsiveness of PAMP or ETI. The resistant cultivar tailored for its defense strategy, WRKY 26 and MYB were down regulated during in-compatible interactions (Fig 9). NAC and MYB genes expressed in both compatible and in compatible interaction of sugarcane and *C. falcatum*. bZIP and

bHLH transcription factor putatively involved in regulation of cell wall integrity was down regulated during in compatible interactions.

#### **Discussion**

Since, sugarcane is a long duration crop, it faces a number of abiotic and biotic stresses that are considered to be highly vulnerable and impact crop growth and productivity. Complex defense response in sugarcane is mainly due to genome complexity which makes different signal perception and transduction networks. Transcription factors (TF) are known for their regulatory role in signalling pathway related to development or disease resistance in sugarcane. TFs activate different pathways directly or indirectly by targeting key genes involved in plant defense responses which are considered to be the host targets by the pathogen (18) (Table 3). Our previous host pathogen interaction studies revealed a lot of pathogenesis related (PR) proteins expressed in compatible and incompatible interactions of sugarcane with C. falcatum, early and prominent induction of chitinase, β-1,3 glucanase and thaumatin-like proteins was documented as a defense and induced defense response against C. falcatum (5,6). Further, accumulation of 3-deoxyanthocyanidin phytoalexins at the pathogen infection site was documented as marker for red rot resistance (19, 20). The differential expression of the chitinase gene in red rot resistant and susceptible sugarcane cultivars was monitored through qRT-PCR (21). Subsequently molecular tools like differential display, subtractive libraries and NGS platforms were deployed to understand differential expression of defense/resistant genes and upregulation of genes during biotrophic phase (7, 8, 9, 22). Further, the involvement of jasmonic acid (JA), ethylene (ET), reactive oxygen species (ROS), phosphoinositide (PI) and calcium (Ca<sup>2+</sup>) signals in the red rot resistance was hypothesized (9, 23).

Distribution of TFs in response to fungal pathogen interaction with different host systems revealed that sugarcane interaction with *C. falcatum* expressed large number of TFs (Table 5). WRKY TFs are alleviating and effective in mediating hormones involved in pathogen-associated molecular patterns (PAMP) triggered immunity, effector triggered immunity or system acquired resistance (SAR) (24). WRKY 100 involved in regulation of biotic and abiotic stress in *C. gloeosporioides* and *M. grisea* (25, 26). WRKY 1 negatively regulates resistance in fruit during infection caused by *C. acutatum* (27). Several reports are made in monocot and dicot plants were WRKY involved in inducing the multiple signal transduction pathway (Table 2), moreover it is suggested that there are around 39 WRKY TFs are available in *Saccharum* spp (18) but during interaction with *C. falcatum* 26 different types of WRKY TFs were found to be expressed under this biotic stress which is considered to be the first consolidated and novel report in sugarcane. Comparative TFs screened in sugarcane challenged with *C. falcatum* using various platforms like transcriptome analysis, SAR priming efficacy, DD RT-PCR and 2D based defense proteome analysis and the results suggested that major TFs were properly screened and consolidated in this study (Table 4). WRKY 51 expressed during sugarcane suspension cells interaction with *C. falcatum* (Table S1). During 2D based defense proteome analysis no WRKY TFs were reported (Table S2).

Another TFs which is equally important is NAC, this NAC TFs plays dual role in triggering innate or hypersensitive response and also ETI. Reports from plant transcription factor databases suggested that

Saccharum spp contains 44 NAC and our transcriptome studies revealed that there are 21 NAC TFs involved during interaction of sugarcane with C. falcatum. There are around 11 NACs were found to be expressed during SAR priming in sugarcane with C. falcatum (28) Over expression of NAC in Oryza sativa (Os NAP) TFs has controlled temperature related stresses and also plays a significant role in increasing resistance against cold stress at vegetative state (29). MYB TFs are considered to be the major and functionally diverse proteins which are primarily involved in protein-protein interactions, DNA binding and protein regulations (18). There are around 36 MYB TFs were reportedly present in Saccharum spp and during interaction with C. falcatum, sugarcane transcriptome contains 44 MYB TFs were found to be expressed (Table 4). MYB TF 1 were down regulated during in compatible interaction with C. falcatum but during compatible interaction MYB 2 were expressed lower (Fig. 8 & 9). bZIP and bHLH TFs majorly reported as stress tolerance factor during *Colletotrichum* infection with concern host system. bZIP in *C.* gloeosporioides and C. coccodes putatively predicted or involved in oxidative stress tolerance, cell wall integrity, host signalling and PAMP triggered immunity (Table 2). C2H2 zinc finger and wing helix repressor DNA binding TFs were expressed during 2D based defense proteome screening which is found to be expressed during compatible interaction in large number and probable TLP TFs were expressed higher during SAR priming but during transcriptome analysis it is found to be expressed less in number (Table 4).

Several reported PR proteins were screened in sugarcane (Table 4), PR 1 a primary protein which is mediated to SA defense response is expressed, PR 2 family (endo-1-3-β-D-glucanase) and PR 5 family (thaumatin – like) associated with pathway regulations, PR 6 (proteinase inhibitor) and PR 9 (Peroxidases) closely associated with JA-ET pathway (28, 30). FAR-1 transcription factors a key component of phytochrome signalling and regulator of ABA signalling was up regulated in both biotic and abiotic stress (31). This FAR-1 was up regulated even after the pathogen reaches necrotrophic phase. Earlier, 85 stress associated NAC genes regulating important NAC plant transcription factor involved in various biological and molecular process were predicted in sugarcane (32). There are several reports that NAC implicated in plant growth, development and flowering (33). R2R3-MYB from MYB family showed that it has been triggered during drought tolerance and alternatively spliced during the plant process (34). ScMYBAS 1 encoding putative transcription factor regulates salinity and water deficit stress (35).

bZIP transcription factor known for its responses during pathogen attack, bZIP genes expressed in velvet leaf during interaction with *C. coccodes* and reported with temporal expression patterns (36). bZIP transcription factor triggers oxidative stress and cell wall integrity during *C. gloeosporioides* infection (37). WRKY gene triggers defense response in strawberry fruit accomplished during interaction with *C. acutaum* (27). WRKY 100 encodes group I WRKY regulates resistance in *Malus domestica* (Apple) during *C. gloeosporioides* infection (25). Tea plant (*Camellia sinensis*) during interaction with *C. gloeosporioides* expresses several transcription factors such as WRKY, MYB75, ARF, NAC and NFY considered as host targets (38). Overall comparative studies suggested that transcriptome of sugarcane and *C. falcatum* interaction brought an encouraging result in identifying the host targets such as WRKY, MYB, NAC and AP2-ERF. Further, it will be established in screening the spatial and temporal expression of sugarcane during biotic stress.

#### Conclusion

Sugarcane TFs are known for its pivotal role in wide range of molecular events such as triggering plant defense, host signalling, regulation of developmental process etc. Here we report the first comprehensive dataset of TFs expressed during biotic stress in sugarcane and the hemi biotrophic pathogen *C. falcatum* interactions. Though *C. falcatum* is a hemi biotrophic pathogen, it always implies successful invasion into host system at biotrophic phase which makes programmed cell death. The transcriptome of sugarcane interaction with *C. falcatum* attributes to regulation of genes at transcription level which brought new insight that TFs are mediated through signal transduction pathway at different growth stages of sugarcane. Hundreds of TFs were triggered/ expressed during biotic stress caused by *C. falcatum* in sugarcane. Among them, WRKY, NAC, MYB, AP2-ERF play a major role in their biological process and the transcriptome studies suggested that TFs gets activated both phases (BNS) can be considered as host targets. To the best of our knowledge this is considered to be the first consolidated report of expressed TFs in a holistic way. This finding unravels new areas of research that allows better understanding of host responses during *C. falcatum* interaction.

#### **Declarations**

# **Acknowledgements**

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

#### Conflict of interest

The authors declare that they have no conflict of interest

#### **Ethical standard**

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 Bio Project PRJNA380333. The accession number of Sequence Read Achieves (SRA) is SRR5384056.

#### References

- 1. Viswanathan R (2021) Sustainable sugarcane cultivation in India through threats of red rot by varietal management. Sugar Tech 23, 239–253 https://doi.org/10.1007/s12355-020-00882-3.
- 2. Viswanathan R (2021) Red rot of sugarcane (*Colletotrichum falcatum* Went) CAB Rev 16, No. 023, doi: 10.1079/PAVSNNR202116023
- 3. Viswanathan R, Malathi P, Sundar AR, Kaverinathan K, Chhabra ML, Parameswari B, Jothi R (2017) Diversity of *Colletotrichum falcatum* population in India: Comparative virulence at two different agroclimatic regions. Int. Sugar J. 119, 966-977
- 4. Viswanathan R, Padmanaban P, Selvakumar R (2020) Emergence of new pathogenic variants in *Colletotrichum falcatum*, stalk infecting Ascomycete in sugarcane: Role of host varieties. Sugar Tech 22, 473-484.
- 5. Viswanathan R, Nandakumar R, Samiyappan R (2003) Involvement of pathogenesis-related proteins in *Pseudomonas* spp. induced systemic resistance against *Colletotrichum falcatum* in sugarcane. J Plant Dis Protect 110, 524-537
- 6. Viswanathan R, Malathi P, Sundar AR, Aarthi S, Prem Kumari SM, Padmanaban P (2005) Differential induction of chitinases and thaumatin like proteins in sugarcane in response to infection by *Colletotrichum falcatum* causing red rot disease. J Plant Dis Protect 112, 417-425.
- 7. Prathima PT, Raveendran M, Kumar KK, Rahul PR, Ganesh Kumar V, Viswanathan R, Ramesh Sundar A, Malathi P, Sudhakar D, Balasubramanian P (2013) Differential regulation of defense-related gene expression in response to red rot pathogen *Colletotrichum falcatum* infection in sugarcane. Appl. Biochem. Biotechnol. 171, 488–503
- 8. Rahul PR, Kumar VG, Viswanathan R, Sundar AR, Malathi P, Prasanth CN, Prathima PT (2016) Defense transcriptome analysis of sugarcane and *Colletotrichum falcatum* interaction using host suspension cells and pathogen elicitor. Sugar Tech 18, 16–28.
- 9. Sathyabhama M, Viswanathan R, Nandakumar M, Malathi P, Sundar AR (2015) Understanding sugarcane defence responses during the initial phase of *Colletotrichum falcatum* pathogenesis by suppression subtractive hybridization (SSH). Physiol. Mol. Plant Pathol. 91, 131-140.
- 10. Viswanathan R, Prasanth CN, Malathi P, Sundar AR (2016) Draft genome sequence of *Colletotrichum falcatum* A prelude on screening of red rot pathogen in sugarcane. J Genomics 4(10): 1−3. doi:10.7150/jgen.13585.
- 11. Prasanth CN, Viswanathan R., Krishna N, Malathi P, Sundar AR. (2017) Unravelling the genetic complexities in gene set of sugarcane red rot pathogen *Colletotrichum falcatum* through transcriptomic approach. Sugar Tech 19: 604-615.
- 12. Prasanth CN, Viswanathan R, Malathi P, Sundar AR (2019) Comparative transcriptome analysis of candidate secretory effector proteins from *Colletotrichum falcatum* infecting sugarcane. Agri Gene 13:100089.
- 13. Prasanth CN, Viswanathan R, Malathi P, Ramesh Sundar A. (2021) Development and characterization of genomic SSR marker for virulent strain specific *Colletotrichum falcatum* infecting sugarcane. 3 Biotech 11: Article number: 20 DOI: 10.1007/s13205-020-02572-z.

- 14. Kumar VG, Viswanathan R, Malathi P, Sundar AR., Prasanth CN, Nanda Kumar M (2020) Identification of differential expressed proteins and establishing a defense proteome of sugarcane in response to *Colletotrichum falcatum* infection. J Plant Pathol 102: 685–702 https://doi.org/10.1007/s42161-020-00577-4
- 15. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35(SUPPL.2) 182–185.
- 16. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28: 27–30.
- 17. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Res 40: D109-D114.
- 18. Javed T, Shabbir R, Ali A, Afzal I, Zaheer U, Gao SJ (2020) Transcription Factors in Plant Stress Responses: Challenges and Potential for Sugarcane Improvement. *Plants 9:* 491 https://doi.org/10.3390/plants9040491
- 19. Viswanathan R, Mohanraj D, Padmanaban P, Alexander KC (1996) Synthesis of phytoalexins in sugarcane in response to infection by *Colletotrichum falcatum* Went. Acta Phytopathol. Entomol. Hung. 31: 229-237.
- 20. Nandakumar M, Malathi P, Sundar AR, Viswanathan R (2021) Host-pathogen interaction in sugarcane and red rot pathogen: Exploring expression of phytoalexin biosynthesis pathway genes. Indian Phytopathol. 74:529–535 DOI: 10.1007/s42360-020-00306-y
- 21. Rahul PR, Ganesh V, Sathyabhama M, Viswanathan R, Sundar AR, Malathi P (2015) Characterization and 3D structure prediction of chitinase induced in sugarcane during pathogenesis of *Colletotrichum falcatum*. J Plant Biochem. Biotech. 24:1-8
- 22. Sathyabhama M, Viswanathan R, Malathi P, Ramesh Sundar A. (2016) Identification of differentially expressed genes in sugarcane during pathogenesis of *Colletotrichum falcatum* by suppression subtractive hybridization. (SSH) Sugar Tech 18: 176–183.
- 23. Viswanathan R, Sathyabhama M, Malathi P, Sundar AR (2016) Transcriptome analysis of host-pathogen interaction between sugarcane and *Colletotrichum falcatum* by Suppression Subtractive Hybridization and Illumina sequencing. Proc Intern Soc Sugar Cane Technol 29: 1639-1644.
- 24. Che F. Hu Y, Vannozzi A, Wu K, Cai H, Qin Y, Mullis A, Lin Z, Zhang L. (2017) The WRKY transcription factor family in model plants and crops. Crit. Rev. Plant. Sci. 36, 311–335.
- 25. Zhang F, Wang F, Yang S, Zhang Y, Xue H, Wang Y, Yan S, Wang Y, Zhang Z, Ma Y. (2019) MdWRKY100 encodes a group I WRKY transcription factor in *Malus domestica* that positively regulates resistance to *Colletotrichum gloeosporioides* infection. Plant Sci 286: 68-77. doi: 10.1016/j.plantsci.2019.06.001.
- 26. Ryu HS, Han M, Lee SK, Cho JI, Ryoo N, Heu S, Lee YH, Bhoo SH, Wang GL, Hahn TR, Jeon JS. (2006) A comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. Plant Cell Rep. 25: 836-847. doi: 10.1007/s00299-006-0138-1.
- 27. Higuera JJ, Garrido-Gala. J, Lekhbou A, Arjona-Girona I, Amil-Ruiz F, Mercado JA, Pliego-Alfaro F, Muñoz-Blanco J, López-Herrera CJ, Caballero JL (2019) The strawberry FaWRKY1 transcription factor

- negatively regulates resistance to *Colletotrichum acutatum* in fruit upon infection. Front. Plant Sci. 10: 480
- 28. Muthumeena M, Ashwin NMR, Sundar AR, Malathi P, Viswanathan R (2013) Expression profiling of transcription factors (TFs) in sugarcane X *Colletotrichum falcatum* interaction. J Plant Biochem Phsiol 22: 286-294 10.1007/s13562-012-0157-7.
- 29. Chen X, Wang Y, Lv B, Li J, Luo L, Lu S, Zhang X, Ma H, Ming F (2014) The NAC family transcription factor OsNAP confers abiotic stress response through the ABA pathway. Plant Cell Physiol 55: 604–619.
- 30. Vidhyasekaran, P (2015) Plant hormone signalling systems in plant innate immunity. Plant Signal Behav 2: 10.1007/978-94-017-9285-1.
- 31. Dharshini S, Mohan C, Vignesh D, Gauri N, Narayan A, Manoj VM, Naveenarani M, Kaur L, Mahadevaiah C, Kumar R, Mintu R, Bakshi R, Appunu C. (2018) Differential gene expression profiling through transcriptome approach of *Saccharum spontaneum* L. under low temperature stress reveals genes potentially involved in cold acclimation. 3 Biotech 8: 195. 10.1007/s13205-018-1194-2.
- 32. Manimekalai R, Narayanan J, Gokul M, Athiappan S, Meena Arun, Gomathi R, R Bakshi R (2016) Genome wide analysis of NAC gene family 'sequences' in sugarcane and its comparative phylogenetic relationship with rice, sorghum, maize and Arabidopsis for prediction of stress associated NAC genes. Agri Gene 3: 10.1016/j.aggene.2016.10.003.
- 33. Kim SY, Kim Y, Seo PJ, Bae M, Yoon H, Park CM (2007) Exploring membrane-associated NAC transcription factors in Arabidopsis: Implications for membrane biology in genome regulation. Nucleic Acids Res 35: 203–213.
- 34. Guo J, Ling H, Ma J, Chen Y, Su Y, Lin Q, Gao S, Wang H, Que Y, Xu L (2017) A sugarcane R2R3-MYB transcription factor gene is alternatively spliced during drought stress. Sci Rep 7: 41922.
- 35. Prabu G, Prasad DT (2012) Functional characterization of sugarcane MYB transcription factor gene promoter (PScMYBAS1) in response to abiotic stresses and hormones. Plant Cell Rep 31: 661–669.
- 36. Dauch AL, Jabaji-Hare SH (2006) Metallothionein and bZIP transcription factor genes from velvetleaf and their differential expression following *Colletotrichum coccodes* infection. Phytopathology 96: 1116-1123.
- 37. Li X, Wu Y, Liu Z, Zhang C (2017) The function and transcriptome analysis of a bZIP transcription factor *Cg*AP1 in *Colletotrichum gloeosporioides*. Microbiol Res 197: 39-48. doi: 10.1016/j.micres.2017.01.006.
- 38. Jeyaraj A, Wang X, Wang S, Liu S, Zhang R, Wu A, Wei C (2019) Identification of regulatory networks of microRNAs and their targets in response to *Colletotrichum gloeosporioides* in tea plant (*Camellia sinensis* L). Front Plant Sci 10: 1096. doi: 10.3389/fpls.2019.01096.
- 39. Jimmy JL, Babu S (2019) Gene network mediated by WRKY13 to regulate resistance against sheath infecting fungi in rice (*Oryza sativa* L.) Plant Sci 280: 269–282
- 40. Li X, Ke Z, Yu X, Liu, Z, Zhang C (2019) Transcription factor Cg Azf1 regulates melanin production, conidial development and infection in *Colletotrichum gloeosporioides*. Antonie Leeuwenhoek 112:

- 1095-1104. https://doi.org/10.1007/s10482-019-01243-1
- 41. Frerigmann H, Piślewska-Bednarek M, Sánchez-Vallet A, Molina A, Glawischnig E, Gigolashvili T, Bednarek P (2016) Regulation of pathogen-triggered tryptophan metabolism in *Arabidopsis thaliana* by MYB transcription factors and indole glucosinolate conversion products. Mol Plant 9: 682-695. doi: 10.1016/j.molp.2016.01.006.
- 42. Amorim LLB, da Fonseca Dos Santos R, Neto JPB, Guida-Santos M, Crovella S, Benko-Iseppon AM (2017) Transcription factors involved in plant resistance to pathogens. Curr Protein Pept Sci 18: 335-351. doi: 10.2174/1389203717666160619185308. PMID: 27323805.
- 43. Sun, Y. Wang, C. Tian (2016) **bZIP transcription factor CgAP1 is essential for oxidative stress tolerance and full virulence of the poplar anthracnose fungus** *Colletotrichum gloeosporioides* Fungal Genet Biol 95: 58-66, DOI 10.1016/j.fgb.2016.08.006.
- 44. Verma S, Gazara RK and Verma PK (2017) Transcription factor repertoire of necrotrophic fungal phytopathogen *Ascochyta rabiei*: Predominance of myb transcription factors as potential regulators of secretome. Front Plant Sci 8: 1037. doi: 10.3389/fpls.2017.01037
- 45. Voitsik, Anna-Maria, Münch, Steffen, Deising, Holger, Voll, Lars. (2013) Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection. BMC Plant Biol 13. 85. 10.1186/1471-2229-13-85.
- 46. Pré M, Atallah M, Champion A, De Vos M, Pieterse CM, Memelink J (2008) The AP2/ERF domain transcription factor ORA59 integrates jasmonic acid and ethylene signals in plant defense. Plant Physiol. 147, 1347-57. doi: 10.1104/pp.108.117523.
- 47. Wang, YC., Hao, XY., Wang, L, Xiao B, Wang XC, Yang YJ. (2016). Diverse *Colletotrichum* species cause anthracnose of tea plants (*Camellia sinensis* (L.) O. Kuntze) in China. Sci Rep 6, 35287 https://doi.org/10.1038/srep35287.

#### **Tables**

Table - 1

Sugarcane and C. falcatum interactive transcriptome assembly statistics and distribution

Description	Trinity	CD-HIT	TGICL	Evigene (Unigene)	Fungi	Sugarcane
Total Number of Transcripts	168960	161701	116422	48935	8540	31065
Total transcriptome length incl. gaps	104785585	101697197	81770315	44326869	5661488	34224178
Total transcriptome length w/o gaps	104785585	101697197	81770315	44326869	5661488	34224178
Avg_transcript_size incl. gaps	620.18	628.921	702.361	905.832	662.938	1101.7
Avg_transcript_size w/o gaps	620.18	628.921	702.361	905.832	662.938	1101.7
Minimum Transcript Length	201	201	201	201	201	201
Maximum Transcript Length	7997	7997	8411	8411	3404	8411
Transcripts > 200 b	168960	161701	116422	48935	8540	31065
Transcripts > 500 b	65053	63519	52596	30960	4467	23479
Transcripts > 1 Kb	27497	27205	24614	15302	1352	13650
Transcript N50	846	869	1024	1250	755	1490

Table 2 – SSR search statistics and repeat statistics

S. No	Table of contents	SSR distribution
1	Total number of Sequences examined	48935
2	Total size of examined sequences (bp)	44326869
3	Total number of identified SSR motifs	8451
4	Total number of SSR containing sequences	7425
5	Total number of Sequence containing more than one SSR repeats	902
6	Total number of compound formation	538
7	Mononucleotide	2501
8	Di nucleotide	1131
9	Tri nucleotide	4056
10	Tetra nucleotide	124
11	Penta nucleotide	162
12	Hexa nucleotide	168

Table - 3

Plant transcription factors expressed during interaction with fungal pathogens reported so far

S. No	Transcription factor	Putative/predicted Role	Fungal organism	References
1	WRKY 100	Regulation of biotic & abiotic stress	Magnaporthe grisea	26
2	WRKY 13	Disease resistance in rice sheath blight and sheath rot	Rhizoctonia solani	39
3	bZIP	Oxidative stress, cell wall integrity	Colletotrichum gloeosporioides	37
4	C2H2 Zinc finger	Involved in melanin production and infection	Colletotrichum gloeosporioides	27
5	WRKY1	Negatively regulates resistance in fruit during infection	Colletotrichum acutatum	27
6	MYB	Indolic glucosinolate biosynthesis	Plectosphaerella cucumerina	41
7	FAR1	Encodes zinc finger protein needed for resistance	Pathogen triggered	42
8	bZIP	Oxidative stress, cell wall integrity	Colletotrichum gloeosporioides	43
9	bZIP	Pathogen response and host signalling	Colletotrichum coccodes	36
10	MYB	Defense response	Ascochyta rabiei	44
11	NAC	Abiotic stress and regulation of developmental process	Colletotrichum graminicola	45
12	WRKY 100	Regulation of biotic and abiotic stress	Colletotrichum gloeosporioides	25
13	AP2-ERF	Plant defense	Botrytis cinerea	46
14	MADS	Disease resistance protein	Colletotrichum camalliae	47
15	ARF, MYB, WRKY,	Wide range of molecular events	Colletotrichum gloeosporioides	37
	NAC, NFY			

Table - 4

Comparative transcription factors screened in sugarcane challenged with *C. falcatum* using various platforms like transcriptome analysis, SAR priming efficacy and 2D based defense protein analysis

S. No	Transcription factor	In planta transcriptome (This study)	TF from SAR screening (Muthu Meena et al. 2013)	TF from 2D Protein analysis (Ganesh et al. 2020)
1	bHLH	16	-	-
2	bZIP	19	12	-
3	MYB	44	3	2
4	Zn2 Cys6	114	-	
5	C2H2 zinc finger	40	-	1
6	Winged helix repressor DNA- binding	32	-	1
7	WRKY	26	5	1
8	NAC	21	11	-
9	TLP	1	10	-
10	Others	114	-	15

Table - 5

Distribution of transcription factor in response to fungal organisms in different host systems

S. No	Organism	bHLH	bZIP	МҮВ	Zn2Cys6	C2H2	Wing	others
1	Neurospora crassa	14	12	36	146	50	41	104
2	Laccaria bicolor	14	11	37	120	112	91	228
3	Fusarium oxysporum lycopersici	16	16	14	322	45	43	112
4	Magnaporthe oryzae	10	17	13	151	41	43	112
5	Mycosphaerella graminicola	11	12	36	201	73	115	175
6	Ustilago maydis	12	7	15	94	26	30	88
7	Puccinia graminis	7	7	7	30	49	39	148
8	Blumeria graminis	7	8	12	34	28	36	99
9	Sclerotina sclerotiorum	9	13	13	119	47	36	194
10	Botrytis cinerea	9	11	15	157	37	43	117
11	Phaespheria nodorum	12	12	13	199	51	53	95
12	Pyrenophora tritici	11	12	13	139	44	42	105
13	Cochilobolpus heterostropus	12	9	11	161	38	44	87
14	Ascochyta rabiei	11	10	16	150	61	28	105
15	Colletotrichum falcatum	16	19	44	144	40	32	114

# **Figures**

#### Top-Hit Species Distribution [plant3\_blast]

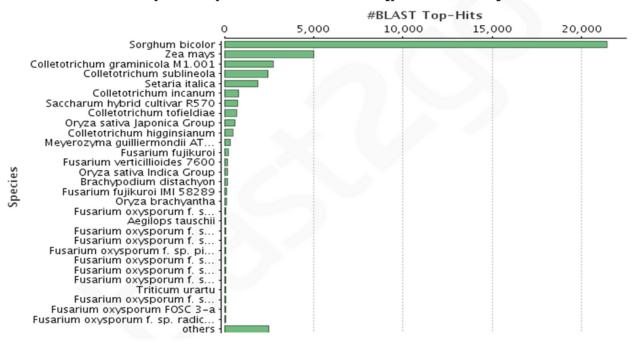
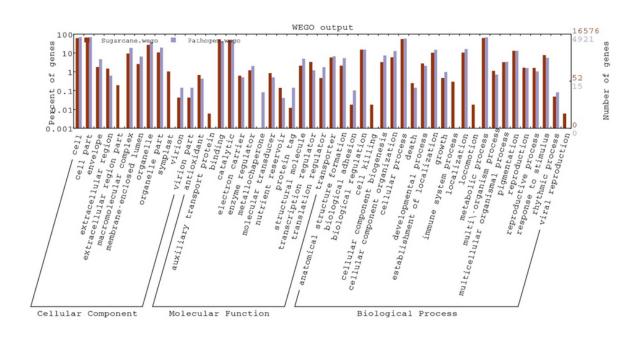


Figure 1

Top species hit distribution



GO analysis from the transcriptome of sugarcane interaction with C. falcatum

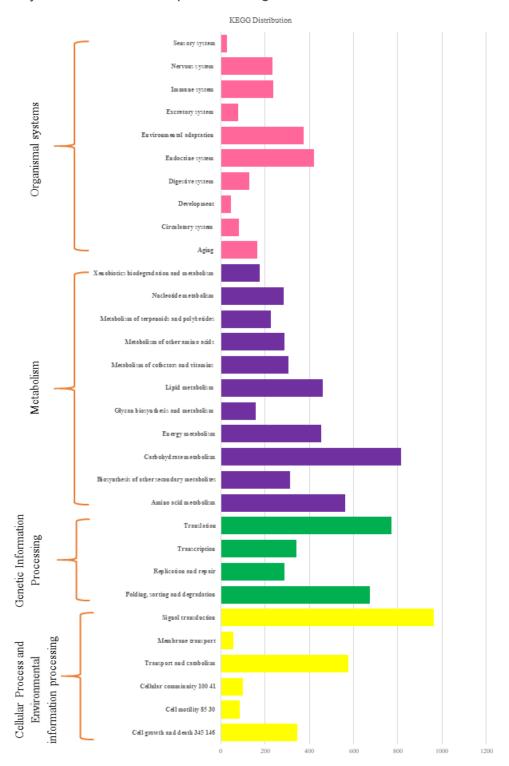


Figure 3

KEGG pathway analysis and Distribution from the transcriptome of sugarcane interaction with C. falcatum

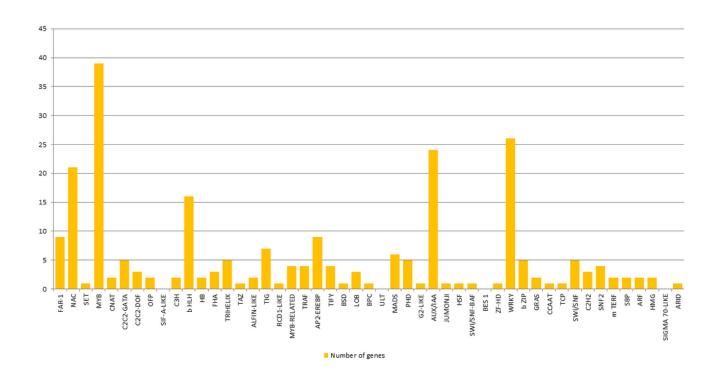
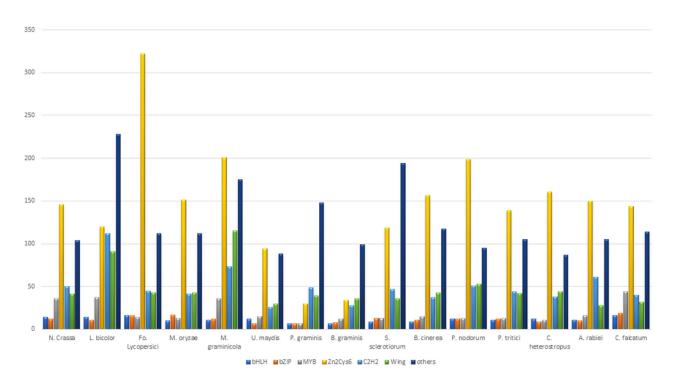


Figure 4

Transcription factor (TF) distribution from the transcriptome of sugarcane interaction with C. falcatum



Distribution of transcription factor in response to fungal organisms in different host systems

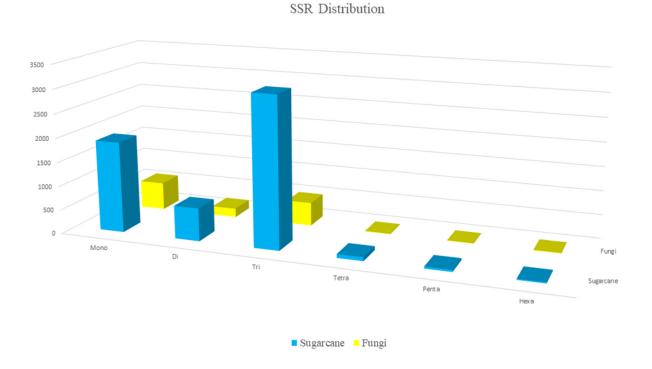
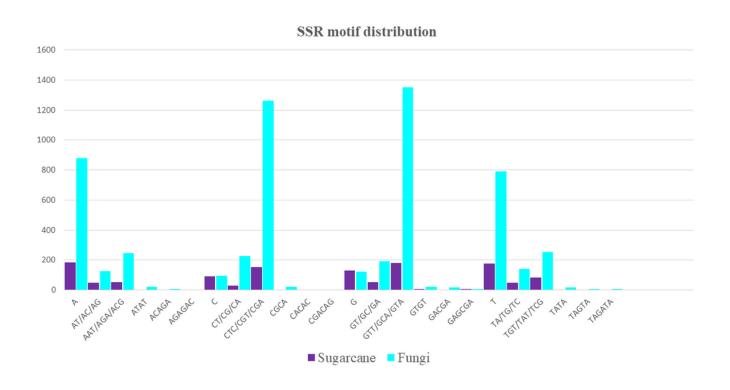


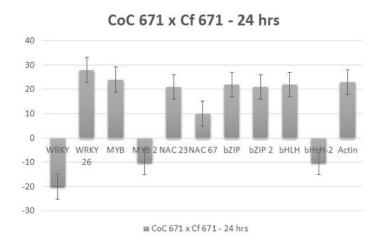
Figure 6

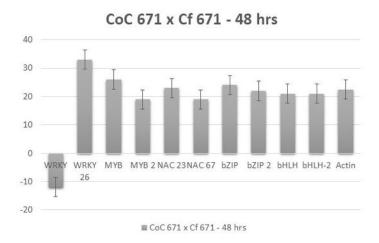
SSR distribution from sugarcane and C. falcatum transcripts



#### Figure 7

List and distribution of different simple sequence repeats identified from sugarcane and C. falcatum transcripts





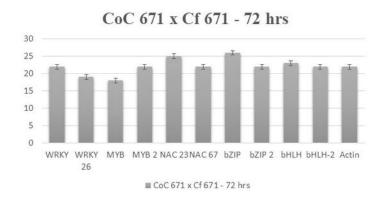
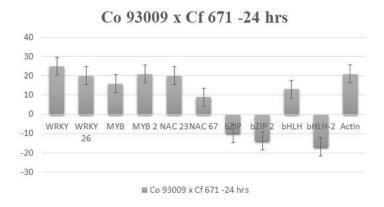
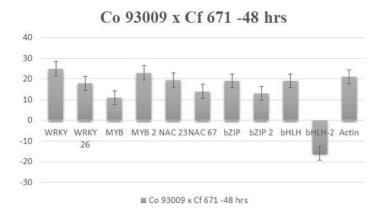


Figure 8

qRT-PCR validation of transcription factor during compatible interactions in sugarcane





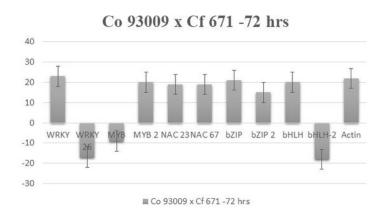


Figure 9

qRT-PCR validation of transcription factor during in compatible interactions in sugarcane

#### **Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

• Supplementaryfiles.docx