University of São Paulo
"Luiz de Queiroz" College of Agriculture

Study of sugarcane metabolism modulation by the plant pathogenic fungus *Sporisorium scitamineum*

Patricia Dayane Carvalho Schaker

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Microbiology

Piracicaba
2016

Patricia Dayane Carvalho Schaker
Bioprocess and Biotechnology Engineering

Study of sugarcane metabolism modulation by the plant pathogenic fungus *Sporisorium scitamineum*

Advisor:
Profa. Dra. CLAUDIA BARROS MONTEIRO-VITORELLO

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Microbiology

Piracicaba
2016

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Schaker, Patricia Dayane Carvalho

Study of sugarcane metabolism modulation by the plant pathogenic fungus *Sporisorium scitamineum* / Patricia Dayane Carvalho Schaker. - - Piracicaba, 2016.

151 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Carvão da cana-de-açúcar 2. Transcriptômica 3. Efetores 4. Metabolômica 5. Interação planta-patógeno I. Título

With love,

To my amazing husband Gabriel,

my parents Jamil and Soili

my brother Felipe and my sister Letícia

ACKNOWLEDGEMENTS

To the University of São Paulo, specially the Superior School of Agriculture "Luiz de Queiroz" – ESALQ for the provision of physical and intellectual infrastructure to develop this project.

To São Paulo Research Foundation for the finantial support to the projects "Study of sugarcane metabolism modulation by the plant pathogenic fungus *Sporisorium scitamineum*" (Grant 2013/25599-2) and "Plant cell localization of candidate effector proteins from *Sporisorium scitamineum*, the sugarcane smut pathogen" (Grant 2016/04429-0).

To National Council for Scientific and Technological Development (CNPq) for the scholarship and finantial support.

To Agricultural Microbiology graduate program, Plant Breeding and Genetics Department and all professors who contributed to my professional training. To the department employees, specially Fernandinho.

To Dr. Claudia Barros Monteiro-Vitorello for the guidance, encouragement, for shraring her knowledgments, experiences and specially her friendship. Our talks always made me think beyond. I'll always be grateful for the opportunity to be in your lab.

To colleagues and friends from Genomics Lab: Nathalia Moraes, Lucas, Giselle, Mariana, Suzane, Daniel, Leandro, Pedro, Marcela, Gustavo, Gian, Sintia, Elaine, Marcos, Ana, Yago, Claudio, Larissa. Especially to Leila, Juliana and Natalia Teixeira for sharing many moments of discussion and friendship; and Lucas for sharing his computational knowledgement and indispensable help in RNAseq data analysis.

To Dra. Alessandra Palhares for the contribution with RNAseq data.

To Dr. Luiz Eduardo Camargo Aranha for the collaboration with qPCR analysis, and especially to his students Mariana, Raphael and Fernanda.

To Dr. Carlos Alberto Labate for collaborating in the metabolomics analyzes (LC-MS), and his team, especially Dr. Thais Regina Cataldi for the willingness to help and teach, and the students Felipe and Ilara for the contributions.

To Dr. Camila Caldana (CTBE) for the collaboration in obtaining and analyzing metabolomics data (GC-MS).

To Dr. Maria Lucia Carneiro Vieira, for assistance in the experiments, and Carla Munhoz for assistance in interpreting the RT-qPCR data.

To Dr. Marie-Anne Van Sluys for the scientific contribution to data interpretation and Sanger sequencing.

To Dr. Silvana Creste for the collaboration with sugarcane plants and data interpretation.

To Dr. S. P. Dinesh-Kumar for receiving me in his laboratory at the University of California, Davis, to carry out effector cell localization experiments, for sharing his expertise and good moments. All the members of Dinesh's Lab, especially Barry, Nagu, Eunsook, JC and Neeraj.

Specially, I thank my parents, Jamil and Soili, my brother Felipe and my sister Leticia, for being my safe haven, for encouraging my choices and believe in my potential. You are examples of hardwork, honesty and wisdom.

Tom my parents in law, Alda and Enio, for the care and the many special moments we shared.

To the most special person I ever met: my husband Gabriel Dequigiovanni, for the companionship and love that made these years pass too fast. For the countless times he saw me better than I am. A new journey starts now, and I can't wait to share more special moments with you!

Finally, thanks God for all wonderful people I met in my way, all the opportunities I had and for being such a blessed woman.

SUMMARY

RESUMO **7**
ABSTRACT **8**
1. INTRODUCTION **9**
1. SUGARCANE AND THE SMUT DISEASE **11**
1.1. STATE OF ART OF SPORISORIUM SCITAMINEUM-SUGARCANE INTERACTION MOLECULAR STUDIES **17**
REFERENCES **20**
2. MODULATION OF SPORISORIUM SCITAMINEUM GENE EXPRESSION PROFILE RELATED TO SUGARCANE COLONIZATION **25**
ABSTRACT **25**
2.1. INTRODUCTION **25**
2.2. MATERIAL AND METHODS **27**
2.2.1. EXPERIMENTAL DESIGN AND RNA EXTRACTION 27
2.2.2. DATA COLLECTION AND ANALYSIS 28
2.3. RESULTS **30**
2.3.1. GENES PREFERENTIALLY EXPRESSED 31
2.3.2. DIFFERENTIALLY EXPRESSED GENES 36
2.3.3. GENES EXPRESSED EXCLUSIVELY DURING INTERACTION 41
2.4. DISCUSSION **42**
REFERENCES **48**
3. PLANT CELL LOCALIZATION OF CANDIDATE EFFECTOR PROTEINS FROM SPORISORIUM SCITAMINEUM, THE SUGARCANE SMUT PATHOGEN **51**
ABSTRACT **51**
3.1. INTRODUCTION **51**
3.2. MATERIAL AND METHODS **53**
3.2.1. EFFECTORS SELECTION 53
3.2.2. PRIMER DESIGN AND AMPLIFICATIONS 54
3.2.3. PLASMIDS AND CLONING PROCEDURES 55
3.2.4. AGROINFILTRATION 55
3.2.5. LIVE-CELL IMAGING BY LASER-SCANNING CONFOCAL MICROSCOPY 56
3.2.6. TOTAL PROTEIN ISOLATION 56
3.2.7. NUCLEAR PROTEIN EXTRACTION 56
3.2.8. IMMUNOBLOT ANALYSES 57
3.3. RESULTS **57**
3.4. DISCUSSION **59**
3.5. PERSPECTIVES **60**
REFERENCES **62**
4. TRANSCRIPTIONAL PROFILE OF SUGARCANE IN SMUT DISEASE: FROM EARLY INFECTION TO TO WHIP DEVELOPMENT **65**
ABSTRACT **65**
4.1. INTRODUCTION **66**
4.2. MATERIAL AND METHODS **68**
4.2.1. ETHICS STATEMENT 68
4.2.2. EXPERIMENTAL DESIGN 68
4.2.3. RNA EXTRACTION, LIBRARIES, AND SEQUENCING 69
4.2.4. PRE-PROCESSING AND MAPPING THE ILLUMINA READS 70
4.2.5. SUGARCANE GENE EXPRESSION ANALYSIS 70
4.2.6. ANNOTATION AND GENE ONTOLOGY ANALYSIS 70
4.2.7. QUANTITATIVE PCR (QPCR) EXPRESSION ANALYSIS 71
4.3. RESULTS AND DISCUSSION **72**
4.3.1. GENERAL ANALYSIS 72
4.3.2. COUNT-BASED DIFFERENTIAL EXPRESSION ANALYSIS OF THE RNASEQ DATA 73
4.3.3. ENRICHMENT ANALYSIS OF GO TERMS 75

4.3.4. HORMONAL IMBALANCE PLAYS A ROLE IN SUGARCANE SMUT DISEASE 79
4.3.5. CARBON DISTRIBUTION IS AFFECTED BY S. SCITAMINEUM COLONIZATION 81
4.3.6. SEQUENCE FEATURES OF RESISTANCE GENE ANALOGS (RGAS) DIFFERENTIALLY EXPRESSED .. 84
4.4. CONCLUSION 86
4.5. DATABASE ACCESSION NUMBER 87
REFERENCES 88
5. ASSESSING THE METABOLIC CLUTTER IN A SUGARCANE-SPORISORIUM SCITAMINEUM COMPATIBLE INTERACTION FROM EARLY INFECTION TO WHIP DEVELOPMENT 94
ABSTRACT 94
5.1. INTRODUCTION 95
5.2. MATERIAL AND METHODS 97
5.2.1. ETHICS STATEMENT 97
5.2.2. EXPERIMENTAL DESIGN 97
5.2.3. QUANTIFICATION OF S. SCITAMINEUM DNA 97
5.2.4. METABOLITES EXTRACTION 98
5.2.5. GC-TOF-MS AND DATA PROCESSING 98
5.2.6. LC-ESI-MS/MS 99
5.2.7. RT-QPCR 100
5.2.8. STARCH STAINING 101
5.2.9. PHENYLALANINE/TYROSINE AMMONIA-LYASE PHYLOGENY 101
5.3. RESULTS 101
5.3.1. PATHOGEN GROWTH WITHIN PLANT TISSUES 101
5.3.2. GC-TOF-MS GENERAL RESULTS 102
5.3.3. EARLY RESPONSE TO SMUT COLONIZATION 104
5.3.4. AMINO ACIDS DIFFERENTIAL ACCUMULATION 104
5.3.5. CARBON PARTITIONING 105
5.3.6. CELL WALL PRECURSORS 107
5.3.7. LC-ESI-MS GENERAL RESULTS 109
5.3.8. SUGARCANE-SMUT SECONDARY METABOLISM PROFILE 110
5.4. DISCUSSION 112
5.5. CONCLUDING REMARKS 116
REFERENCES 118
6. FINAL CONSIDERATIONS 125

RESUMO

Estudo da modulação do metabolismo da cana-de-açúcar pelo fungo fitopatogênico *Sporisorium scitamineum*

Esta tese apresenta uma compreensão mais aprofundada da interação entre o fungo patogênico *Sporisorium scitamineum* e a cana-de-açúcar, doença conhecida como "carvão da cana". O desenvolvimento de uma longa estrutura similar a um "chicote" a partir do meristema de plantas infectadas é a principal característica da doença, permitindo a efetiva dispersão dos teliósporos no campo. As plantas doentes apresentam um teor reduzido de sacarose e qualidade do sumo, levando a perdas econômicas consideráveis. No primeiro capítulo, o perfil de expressão gênica do patógeno durante o seu desenvolvimento *in planta* – nos primeiros momentos da infecção e após a emissão do chicote - e *in vitro* foi avaliado utilizando a técnica RNA-Seq. Foram analisados os genes preferencialmente expressos em cada condição, diferencialmente expressos em relação ao crescimento em meio de cultura, ou expressos apenas durante a interação. Os resultados permitiram a elaboração de hipóteses sobre os mecanismos de patogenicidade, sobre os genes candidatos a efetores ativos e a identificação de agrupamentos de genes expressos apenas durante a interação. No segundo capítulo, para determinar o compartimento celular alvo de alguns dos efetores candidatos e estabelecer um protocolo viável para o estudo de proteínas de *S. scitamineum* foi utilizada a técnica de expressão transiente. Os quatro genes mais expressos durante os momentos iniciais da interação que fazem parte do secretoma do fungo foram fusionados ao gene que codifica a proteína verde fluorescente (Citrina) e expressos em *Nicotiana benthamiana*. Os resultados de microscopia confocal e *westernblots* indicaram um acúmulo de cada uma das proteínas candidatas na membrana, citosol e/ou núcleo, além da ocorrência de modificações pós-traducionais. Esses dados oferecem novas oportunidades de estudo para a identificação de proteínas vegetais que interagem com tais efetores. No terceiro capítulo, as respostas transcricionais da cana-de-açúcar nos primeiros momentos de uma interação compatível e após o desenvolvimento do chicote foram analisadas utilizando novamente os dados obtidos a partir do *dual* RNAseq cana-carvão. Entre as principais respostas da cana destacou-se um aumento da expressão de genes que codificam fatores de transcrição do tipo MADS, indicando que o desenvolvimento do chicote pode usar uma rota semelhante à do florescimento, cuja sinalização parece iniciar logo nos primeiros momentos de colonização. Além disso, o desenvolvimento do chicote é acompanhado pelo aumento da transcrição de genes envolvidos em vias energéticas, e vias de síntese e sinalização hormonal. Genes que codificam para RGAs foram diferencialmente expressos e podem estar relacionados ao reconhecimento de efetores. No quarto capítulo, foi avaliado o perfil metabólico da cana-de-açúcar durante a progressão da doença, confirmando que no meristema de plantas infectadas ocorre um aumento da alocação de carbono em vias energéticas, além da regulação de vários aminoácidos e mudanças em relação à composição da parede celular em resposta ao desenvolvimento do chicote. A abordagem metabólica também permitiu a identificação de uma provável micotoxina derivada de *S. scitamineum*. Os resultados obtidos neste estudo contribuíram para aumentar a compreensão da interação entre *S. scitamineum* e a cana-de-açúcar que se caracteriza pela alta complexidade e especialização ao hospedeiro, e poderão ser utilizados de forma a auxiliar a caracterização de variedades resistentes e contribuir para o melhoramento da cana-de-açúcar com resistência ao carvão.

Palavras-chave: 1. Cana-de-açúcar 2. Carvão 3. Transcriptoma 4. Metabolômica 5. Expressão transiente 6. Efetores

ABSTRACT

Study of sugarcane metabolism modulation by the plant pathogenic fungus Sporisorium scitamineum

This thesis presents a more in-depth understanding of the interaction between the pathogenic fungus *Sporisorium scitamineum* and sugarcane, a disease known as "cane smut". The development of a long structure like a "whip" from the meristem of infected plants is the main characteristic of the disease, allowing the effective dispersion of teliospores in the field. Infected plants have a reduced sucrose content and juice quality, leading to considerable economic losses. In the first chapter, the gene expression profile of the pathogen during its development *in planta* - in the first moments of infection and after the emission of the whip - and *in vitro* was evaluated using the RNAseq technique. Were analyzed genes preferentially expressed in each condition, differentially expressed in comparison to its growth *in vitro*, and expressed only during interaction. The results allowed the identification of some potential pathogenicity mechanisms, active effectors and gene clusters expressed only during interaction. In the second chapter, the transient expression technique was used to determine the target cell compartment of some of the candidate effectors and to establish a viable protocol for the study of *S. scitamineum* proteins. The four putatively secreted genes most expressed during the initial moments of the interaction were fused to the gene encoding the fluorescent green protein (Citrine) and expressed in *Nicotiana benthamiana*. The results of confocal microscopy and *westernblots* indicated an accumulation of each candidate protein in the membrane, cytosol and/or nucleus, in addition to the occurrence of post-translational modifications. These data offer new study opportunities for the identification of plant proteins that interact with such effectors. In the third chapter, the transcriptional responses of sugarcane in the first moments of a compatible interaction and after the development of the whip were analyzed using again the data obtained from the dual RNAseq cane-smut. Among the main responses, was identified an increase in MADS-type transcription factors expression, indicating that the whip development may use a route similar to flowering, whose signaling seems to start as early as the colonization. In addition, whip development is accompanied by increased transcription of genes involved in energetic pathways, and hormones synthesis and signaling pathways. Genes encoding RGAs were differentially expressed and may be related to pathogen effector's recognition. In the fourth chapter, the metabolic profile of sugarcane was evaluated during disease progression, confirming that in the meristem of infected plants carbon allocation is channeled to energetic pathways, besides the regulation of several amino acids and changes in plant cell composition in response to whip development. Metabolomics approach also allowed the identification of a probable mycotoxin derived from *S. scitamineum*. The results obtained in this study contributed to increase the understanding of the interaction between *S. scitamineum* and sugarcane that is characterized by high complexity and specialization to the host, and can be used in a way to help the characterization of resistant varieties and contribute to the improvement of sugarcane with resistance to smut.

Keywords: 1. Sugarcane 2. Smut 3. Transcriptome 4. Metabolomics 5. Transient expression 6. Effectors

1. INTRODUCTION

Sugarcane is one of the most economically valuable crop worldwide, used as the main raw material to sugar and ethanol production. However, decrease in productivity due to biotic and abiotic stresses is a shared concern among producers. Regardless being a very rustic crop, sugarcane hosts several important pathogens that threatens every year productivity because the appearance of new pathogenic races. Sugarcane smut is one of the most harmful disease to the culture, causing losses in all sugarcane-producing countries due to a reduction in sugar content and juice quality. The disease is caused by the biotrophic basidiomycete *Sporisorium scitamineum*. The infection initiates with teliospore germination originating haploid cells of opposite mating types, which may combine to form the infective dikaryotic hyphae and colonize sugarcane tissues leading to teliospores differentiation. Disease cycle ends with the development of a structure like a whip, where billion of teliospores are formed and can easily be spread in the field by wind, achieving germinating buds and restarting the infection cycle.

This work brings new pieces to solve the "puzzle" of this unique interaction using omics approaches to evaluate both fungus and plant responses. The thesis was built on the hypothesis that *S. scitamineum* activates the expression of genes related to pathogenicity including those encoding a set of uncharacterized secreted proteins that may act as effectors inside different compartments of plant cells; whereas, sugarcane answers at transcriptional and metabolic levels to deal with *S. scitamineum* colonization. The understanding of this molecular cross-talking may bring new clues about genes and pathways contributing to plant susceptibility and disease symptoms development. To validate these hypotheses, RNAseq technique was used to evaluate *S. scitamineum* transcriptional profiles in early stages of the interaction and after whip development compared to its gene expression profile in culture medium growth, allowing the determination of pathogen set of genes preferentially expressed in each condition and genes expressed only during interaction. Predictions regarding effector proteins were depicted in a second step, by assessing plant cell compartment targeted by *S. scitamineum* candidates using *Nicotiana benthamiana* transient expression, as a first attempt to determine hereafter plant receptors. To access sugarcane responses to smut differential transcriptional profile of the susceptible sugarcane variety "RB925345" in the early stages of smut disease and after whip development was determined. Because of a strong metabolic change of sugarcane symptomatic plants, the metabolic profile of plants during the disease progression was compared to plants of normal growth. The results of this work provide valuable information about the sugarcane smut

disease, and represent a starting point for further research aiming the understanding of resistance mechanisms and factors involved in pathogen recognition.

1. SUGARCANE AND THE SMUT DISEASE

Sugarcane (*Saccharum* spp.) is a crop of vegetative propagation originated in the Southeast Asia (Daniels and Roach, 1987), currently being cultivated in more than 100 tropical and subtropical countries and representing approximately 0.5% of the total area used for agriculture in the world (FAOSTAT, 2013). Sugarcane is responsible for 70% of world's sugar production, and among the crops with bioenergy potential, it is the one that stands out because presents several favorable attributes, such as high sucrose yield, rapid growth and survival to adverse conditions (Waclawovsky *et al.*, 2010). The output to input ratio (i.e. the ratio of the energy contained in a given volume of ethanol divided by the fossil energy required for its production) of sugarcane first generation ethanol production is around 8.2 to 10, compared to 1.3 in maize (Goldemberg, 2008). Given the increased demand for renewable fuels in recent decades to meet the global agendas of providing energy security and confronting climate change, sugarcane has been also recognized by its potential for second generation ethanol production, where lignocellulosic biomass is hydrolyzed to release fermentable sugar monomers (Botha and Moore, 2014)

The cane-plant is a semi-perennial crop harvested from 12 to 18 months after planting in tropical conditions, while the cane-ratoon is harvested at intervals equal to or smaller, and can be grown preferentially for up to five cycles (Scortecci *et al.*, 2012). In Brazil, the cultivated area destined to sugarcane activity in 2015/16 was approximately 10,870,647 hectares distributed in all producing regions, yielding the harvest of 666,824 thousand tons. São Paulo state holds the first position in sugarcane production (52.7 %). In the current year, it was registered a production of 30,232 thousand m^3 of ethanol and 33,827 thousand tons of sugar (UNICA, 2016). Additionally, sugarcane is used as raw material for animal feed, production of cachaça, bioplastics and co-generation of energy.

Belonging to the Poaceae family, along with other crops of great economic importance such as maize, wheat, barley, rice, rye and oats, modern sugarcane varieties are complex interspecific hybrids of the genus *Saccharum* (Amalraj and Balasundaram, 2006; D'Hont, 2005). The first artificial interspecific hybrids of sugarcane were obtained in Java, India, 1885, encouraged by the onset of the *sereh* disease. The species of *Saccharum* that are most relevant to the development of modern varieties of sugarcane are *S. spontaneum* and *S. officinarum*, which combine the hardiness and disease resistance to good physiological qualities and the high sugar content, respectively (Cheavegatti-Gianotto *et al.*, 2011). The recovery of high sugar content was

carried out by repeated backcrosses with *S. officinarum*, in a process called "nobilization" (Bremer, 1961).

S. officinarum (2n = 80), known as "*noble cane*", accumulate high concentrations of sucrose, but has low resistance to diseases. *S. officinarum* is not a simple polyploid (x = 10 chromosomes) but a complex hybrid of different species with autopolyploidy and allopolyploidy (D'Hont *et al.*, 1996; Sreenivasan, 1987) and probably originated from the introgression of *S. spontaneum*, *S. arundinaceus*, *S. robustum* and *Miscanthus sinensis* (Daniels and Roach, 1987).

S. spontaneum is considered the most primitive species and features x = 8 chromosomes (Sreenivasan, 1987), with variation in the number of copies of each chromosome (2n = 40-128). It is characterized as a more adaptable and vigorous species that grows in a variety of environments, and presents resistance to diseases (D'Hont *et al.*, 1996). It is also recognized by its contribution to the improvement in sugarcane hardness, tillering and ratooning ability (Sreenivasan, 1987).

Interspecific hybridizations allowed a leap in cane breeding improvement, solving many of the problems related to diseases, together with other benefits such as increased productivity (Roach, 1972). However, sugarcane became one of the most genetically complex crop studied to date (Piperidis *et al.*, 2010). Genomic *in situ hybridization* technique showed that in modern varieties approximately 70 to 80% of chromosomes are derived from *S. officinarum* and 10 to 23% from *S. spontaneum*, the remaining are recombinant chromosomes of both species (D'Hont *et al.*, 1996; Piperidis *et al.*, 2010). Sugarcane also presents high degree of aneuploidy, 2n = 108 to 118 for the modern cultivars and 2n = 112–119 for the breeding clones (Piperidis *et al.*, 2010), and its genome size ("R570", 2n=115) is estimated as 10,000 Mpb (D'Hont and Glaszmann, 2001).

Attempts to understand sugarcane genome and the set of genes involved in important characteristics, such as sugar accumulation and disease resistance, are been achieved by several strategies. Development of ESTs (*Expressed Sequence Tags*) technique facilitated the access of the genetic information of this very complex genome. As a first step in depicting the sugarcane genome, the ONSA consortium ("*Organization for Nucleotide and Sequencing Analysis*") launched the *Sugarcane Expressed Sequence Tag* project (SUCEST). The 26 SUCEST cDNA libraries were constructed from different sugarcane organs and tissues sampled at various developmental stages, producing 237,954 high-quality ESTs that were assembled into 43,141 putative transcripts, referred as the *Sugarcane Assembled Sequences* (SAS) (Vettore *et al.*, 2001). These transcripts represent 33,620 expressed sugarcane genes and were grouped into 18 broad categories of biological roles and constitutes an important resource for the genomics of sugarcane and related species (Vettore *et al.*, 2003).

More recently, a new set of sugarcane transcripts was analyzed in a *de novo* assembly approach of RNAseq using six sugarcane genotypes, releasing 72,269 sugarcane unigenes. Translated sequences were similar to more than 28,788 sorghum proteins and included a set of 5,272 unigenes not present in the EST database (Cardoso-Silva *et al.*, 2014). Although sugarcane genome sequencing initiatives are being performed by several groups, so far its complete sequence is not yet available. As part of the sequencing initiative to determine sugarcane genome structure and function 317 chiefly euchromatic inserts cloned in BACs were sequenced, leading to 1,400 manually-annotated protein-coding genes, including some related to sucrose and starch metabolism pathway (de Setta *et al.*, 2014). Nowadays, NCBI database hosts more than 280,000 ESTs, 38,000 DNA and RNA sequences, 83,000 GSS (*Genome Survey Sequences*), 43,000 genomic and cDNA clones, 18 genome sequencing projects and 259 SRA archives (high-throughput DNA and RNA *Sequence Read Archive*) related to sugarcane.

This accumulated information allows researchers to better understand the plant responses to biotic and abiotic stresses by creating hypotheses that may be tested in experimental designs. Among the main factors that can influence the yield of sugarcane are diseases caused by bacteria, fungi and viruses. More than 100 pathogens have been described as causing disease in sugarcane (Rott, 2000). *Sporisorium scitamineum* (Syd.) M. Piepenbr., M. Stoll & Oberw is the causal agent of sugarcane smut, one of the major diseases that affect the culture worldwide (Sundar *et al.*, 2015), featuring a biotrophic lifestyle with inter- or intracellular growth (Carvalho *et al.*, 2016; Stoll *et al.*, 2005). The presence of the fungus in the plant leads to the development of a long structure like a "whip" in susceptible varieties at the apex of the apical or lateral meristems (Figure 1A, B; Sundar *et al.*, 2012). This structure comprises a central region formed by parenchymal and vascular tissues of the host plant, around of which the teliospores are produced (Singh *et al.*, 2004).

Infected plants may also have a highly significant decrease in stalks height and diameter, reduced weight and juice recovery, smaller and narrowed leaves and formation of galls on the sides of the stem buds (Sundar *et al.*, 2012). Additionally, it was described a reduction of sucrose (% Brix), % Pol, % Purity and % Fiber and increased % reducing sugars in juice of infected plants (Wada *et al.*, 2016).

Figure 1. Sugarcane smut disease symptoms in RB925345 variety after whip development. (A-B) Whip development. (C) Reduction in culm diameter. (D) Tillering. (E) and (F) Galls on the sides of the stem buds.

Sugarcane smut infection initiates with teliospores (2n) germination originating four haploid (n) sporidal cells of compatible mating types (+ and -). The fusion of these compatible cells originates the infective dikaryotic hypha (n + n) (Bakkeren *et al.*, 2008; Waller, 1970). The genes of mating type locus *a* encode to a membrane receptor and a pheromone that are compatible with the membrane receptor and pheromone of the opposite mating type. *S. scitamineum* is able to form appressorium mainly upon undifferentiated epidermal tissues, such as meristem buds on germination, which is the primary pathogen entry point (Waller, 1970).

Appressorium development is dependent on the formation of a heterodimeric transcription factor composed of subunits encoded by different alleles of the *b* locus. Recently, Yan and coworkers (2016) obtained mutants of a mating type "+" strain expressing genes from locus *b* from a mating type "-" strain. The mutant cells were able to promote filamentous growth without mating, suggesting that a functional heterodimer codified by the *b* locus is sufficient to maintain the filamentous growth. However, plants inoculated with the mutant strain do not developed whips, suggesting that the response mediated by the compatibility of pheromones in *a* locus is required for pathogenicity (Yan *et al.*, 2016).

The smut disease culminates with teliospores differentiation as a result of karyogamy and hyphal fragmentation of at the base of whip (Marques *et al.*, 2016). Millions of mature teliospore can be easily dispersed by wind, spreading the disease in the field. In more susceptible varieties whips can be observed up to 6 weeks after the infection (Singh *et al.*, 2004) and the losses caused by the disease can achieve up to 30% of productivity reduction (Magarey *et al.*, 2011).

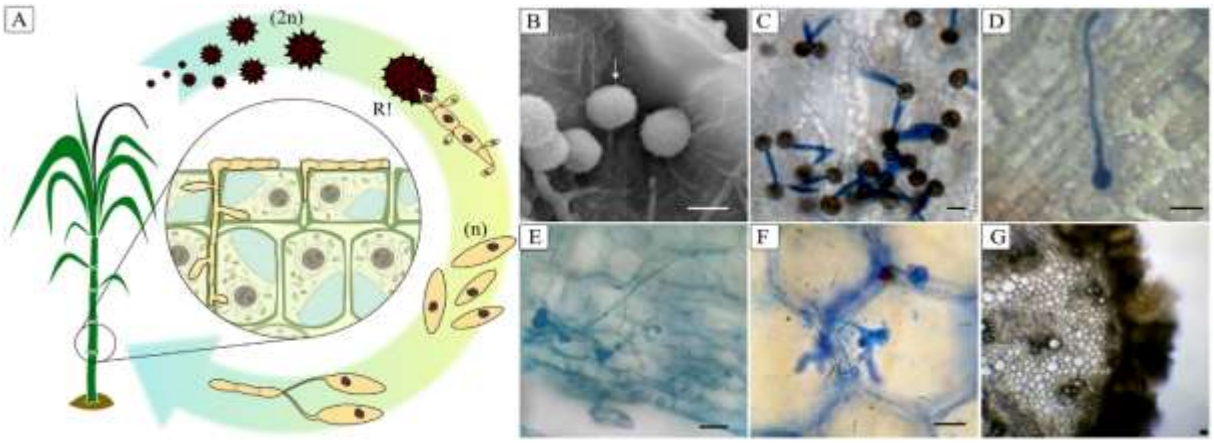


Figure 2. Life cycle of *Sporisorium scitamineum*. (A) Developmental stages in the S. scitamineum life cycle: diploid teliospores (2n); haploid yeast-like sporidia (n) after meiosis (R!); hyphal fusion. (B) Scanning electron micrograph of spores adhered to sugarcane bud surface. (C) Germination of spores on bud scale epidermis and tube-like promycelium formation at 6 hai (hours after inoculation); photomicrograph of tube-like promycelium stained with lactophenol-cotton. (D) Photomicrograph of appressorium formation 48 hai stained with lactophenol-cotton blue; (E) Photomicrograph of S. scitamineum growth on parenchyma cells of bud tissue observed at 120 hr stained with lactophenol-cotton blue. (F) Photomicrograph of S. scitamineum intracellular growth on parenchyma cells of white whip portion; stained with lactophenol-cotton blue. (G) Photomicrograph of black whip portion showing the mature spore liberation. Scale bar = 5 µm. Source: Taniguti et al., 2015.

Sugarcane smut occurs in all sugarcane-producing countries, except Papua New Guinea, and has produced epidemics worldwide. With its outbreak in cane fields, several strategies were adopted to prevent smut dissemination or to reduce the effects of the disease. Among them are the screening for resistant varieties (Croft *et al.*, 2008; Lemma *et al.*, 2015; Nalawade *et al.*, 2012; Singh *et al.*, 2005), the use of fungicides in preplant or post-plant (Agboire *et al.*, 2003; Bhuiyan *et*

al., 2012; Bhuiyan *et al.*, 2015; Olufolaji, 1993), the hot water treatment of planting setts (Gupta, 1978; Schenck, 2003), the crop rotation and soil flooding (Abdou *et al.*, 1990), and rouging of smutted stools (Antoine, 1961).

Screening for smut resistance in breeding programs usually takes place at the first stages of selection to avoid carrying large numbers of clones to advanced stages of selection. To test for resistance, buds are subjected to artificial inoculation and then grown in nurseries containing several different varieties of sugarcane. The method of inoculation usually adopted is to deep buds in aqueous suspensions of spores. The percentage of whips among the following season's stalks is then taken as a measure of the susceptibility of a variety to the disease (Elston and Simmonds, 1988). Data on smutted stool begins to be collected after six weeks and continues until ten months after planting (Lemma *et al.*, 2015). Resistance to smut is based in a scale that ranges from 0 (Immune) to 9 (very highly susceptible) (Latiza *et al.*, 1980). However, varieties classified as resistant by this approach may suddenly develop whips, according to environmental conditions, since the pathogen is able to colonize sugarcane tissues of resistant varieties (Carvalho *et al.*, 2016).

An early and precise diagnosis of smut susceptible clones in breeding programs, as well as plants suspected to be contaminated in fields, is an important component of a successful management of sugarcane smut. Early detection can be very useful in reducing the costs of the quarantine process. Nowadays, smut detection before whip emission is possible using molecular techniques and microscopy (Albert and Schenck, 1996; Bueno, 2010; Kavitha *et al.*, 2014; Lloyd and Naidoo, 1983; Lloyd and Pillay, 1980; Singh *et al.*, 2004; Sinha *et al.*, 1982). PCR assay is more efficient for smut detection than microscopy as observed in Singh and coworkers (2004) and Kavitha and coworkers (2014) studies. Additionally, TaqMan qPCR has a higher sensitivity compared to conventional PCR (Su *et al.*, 2013).

Sugarcane resistance to smut was demonstrated as an inheritable feature, however the genetic determinants are still unknown (Hector *et al.*, 1995; Lloyd and Naidoo, 1983; McNeil *et al.*, 2011; Sundar *et al.*, 2015). Interesting is the fact that although *S. spontaneum* is recognized by its genetically contribution to disease resistance in modern sugarcane, it has recently been demonstrated its interaction with *S. scitamineum* (Jose *et al.*, 2016). The authors showed that *S. spontaneum* uninfected plants are taller and develop inflorescence. The infection does not occur in a systemic way, unlike what happens in sugarcane, which could explain the formation of internode bulges characteristic of the *S. scitamineum* infection in *S. spontaneum*. Furthermore, it is hypothesized that the bulges caused by a confined pathogen growth may prevent the transport of nutrients necessary for inflorescence development (Jose *et al.*, 2016).

1.1. State of art of *Sporisorium scitamineum*-sugarcane interaction molecular studies

The plant defense against pathogens may be classified broadly into pre- and post-formed barriers. Regarding to sugarcane smut, the biochemical composition and the number of trichrome in buds are probably the main preformed defenses (Waller, 1970). If the resistance is based only on the morphology of the buds, sprouting makes the plant more susceptible to pathogen entry. The access of the spores is easier since the protective scales are detached increasing the area for hyphal penetration (Waller, 1970). Inhibition of teliospore germination by flavonoids presents in buds was described (Lloyd and Pillay, 1980), and the relationship between bud glycosidic substances and resistance to pathogen was reported (Lloyd and Naidoo, 1983).

After pathogen perception, post-formed barriers are activated. In the *S. scitamineum* x sugarcane interaction many post-formed mechanisms have been described, among them production of glycoproteins that impairs teliospore germination by preventing polarization through inhibition of germ tube protrusion (Fontaniella *et al.*, 2002; Millanes *et al.*, 2005), accumulation of free or combined polyamines in plant tissues (Legaz *et al.*, 1998; Piñon *et al.*, 1999), increased levels of salicylic acid (Borrás-Hidalgo *et al.*, 2005), lignification of infected tissues (Santiago *et al.*, 2009; Santiago *et al.*, 2010; Santiago *et al.*, 2012), and increased activity of PR (*Pathogenesis Related*) proteins, such as phenylalanine ammonia lyase, peroxidase, esterase and chitinase (Esh, 2014).

Efforts to elucidate the overall defense response was performed using gene expression analysis (Borrás-Hidalgo *et al.*, 2005; Huang *et al.*, 2015; LaO *et al.*, 2008; Que *et al.*, 2014; Thokoane and Rutherford, 2001; Wu *et al.*, 2013) and protein differential accumulation (Barnabas *et al.*, 2016; Que *et al.*, 2011), mostly to describe significant changes related to resistance in early interaction. Thokoane & Rutherford (2001) used cDNA-AFLP technique (cDNA-*Amplified Fragment Length Polymorphism*) to determine genes expressed only in the resistant variety "N52/219" after inoculation when compared to the susceptible variety "Co301". Among the polymorphisms were identified proteins like-kinase and a Pto kinase receptor (*S-receptor-like kinase* - RLK) involved in the recognition of chitin. The same technique was used to demonstrate that two months after infection the resistant variety "JaS-44" has among differentially expressed genes an NBS-LRR-like (*Nucleotide-Binding Site - leucine-Rich Repeat*), and genes encoding enzymes of the phenylpropanoid pathway, such as phenylalanine ammonia lyase and cinnamyl-alcohol dehydrogenase; and ethylene biosynthesis, such as ACC oxidase (Borrás-Hidalgo *et al.*, 2005). Later, LaO and coworkers (2008) assessed differential gene expression in the early stages of the interaction (24 and 72 hours) using cDNA-AFLP in susceptible and resistant varieties, showing that among early events related to resistance occurs an increased expression of peroxidase,

involved in the oxidative burst, which in turn may contribute to increased expression of cinnamoyl alcohol dehydrogenase, related to lignification of tissues. The results also suggest a relationship between resistance and an increase ethylene and auxin signaling along with increased expression of a NBS-LRR protein, homologous to the gene *rgaS* from *Hordeum vulgare* (LaO *et al.*, 2008). More recently, the cDNA-AFLP technique was used again in the resistant variety "NCo376", and were identified 91 polymorphic bands and 45 with increased expression after 12 to 72 hours of inoculation. The sequenced bands were functionally classified as related to defense, energy metabolism, transport, signal transduction, nucleic acid metabolism, transcription and protein synthesis (You-Xiong *et al.*, 2011).

Heinze and coworkers (2001) used the SSH (*Suppression Subtractive Hybridization*) technique to detect kinase receptors and genes involved in flavonoid metabolism in the resistant variety "N52/219". Wu and coworkers (2013) used the tag-seq Solexa technique to identify differentially expressed genes in contrasting varieties for smut resistance 12, 24, 36, 48 and 72 hours after inoculation. The authors suggested the use of genes encoding MAP kinases as molecular markers for resistance (Wu *et al.*, 2013). Later, Que and coworkers (2014) used RNA-Seq technique to obtain the transcript profile of a resistant variety (Yacheng05-179) and susceptible ("ROC"22) at 24, 48 and 120 hours after inoculation, revealing an earlier induction of gene expression in the resistant variety (24 and 48 h) compared to the susceptible (120 h). Among the genes differentially expressed in resistant variety were detected some related phytohormone jasmonate (JAZ and MYC2), flavonoid biosynthetic pathway (PAL, cinnamate 4-hydroxylase, 4-coumarate CoA ligase), PR proteins (glucanase, chitinase and catalase), fortification of cell wall enzymes (syntaxin, hydroxyproline-rich protein, and a gene related to wax) and several transcription factors (MYB, WRKY and ERF) (Que *et al.*, 2014).

In addition to the transcriptional profile, studies have also analyzed the enzymatic activity and the differential accumulation of sugarcane proteins during the infection by *S. scitamineum*. The PR protein β-1,3-glucanase, involved in the breakdown of the fungal cell wall, presents an increased activity in the first 12 hours after inoculation, reaching a maximum at 24 hours in the resistant variety "Yacheng05-179", whereas in susceptible "Liucheng03-182" occurs a reduction in its activity at 12 hpi, indicating a positive correlation between β-1,3-glucanase activity and resistance to smut (Ya-chun Su *et al.*, 2013). Two additional studies used SDS-PAGE technique to detect proteins involved in resistance (Que *et al.*, 2011) and the development of disease symptoms (Barnabas *et al.*, 2016). The first group obtained the +1 leaf protein profiles of susceptible and resistant varieties after whip emission, showing differential protein accumulation related to photosynthesis, signal transduction and disease resistance (Que *et al.*, 2011). More

recently, protein profile of a susceptible variety was obtained from the apical meristem region after whip development. This study revealed differential accumulation pattern of 53 proteins, most of them are part of the phenylpropanoids, carbohydrate and amino acid pathways (Barnabas *et al.*, 2016). Early interaction proteome profiles from resistant and susceptible varieties confirmed some gene regulation results, such as increased levels of beta-1,3-glucanase, peroxidase, ethylene and gibberellic acid pathways, phenylpropanoid metabolism and PRs, such as PR1, PR2, PR5 and PR14 (Su *et al.*, 2016).

Considering the importance of smut disease in world's economic scenario, and the lack of studies aiming to describe host responses in compatible interactions and linking gene expression to metabolic responses, this thesis was defined based on the use of complementary technologies, i. e. RNAseq, transient expression and metabolomics, to provide new insights about this unique interaction. In Chapter 1 are presented the results of differential expression analysis of *S. scitamineum* growing *in vitro* and *in vivo* conditions, including a set of candidate effector proteins; in Chapter 2 the transient expression technique was used to determine plant compartment targeted by fungal putative effectors most expressed in early infection; Chapter 3 focused on the differential expression of sugarcane genes early after inoculation and after whip development; and in Chapter 4 metabolome responses of sugarcane plants during the disease development were assessed, and includes the development of primers to *S. scitamineum* quantification *in planta* using qPCR technique.

REFERENCES

Abdou, Y.A., Moursy, M.A., Abdel-Fattah, M.N.D. and Mansour, I.M. (1990) Effects of temperature and certain cultural practices on longevity of teliospores of *Ustilago scitaminea*. *Bull. Fac. Agric. Univ. Cairo* **41**, 511–520.

Agboire, S., Wada, A.C. and Ishaq, M.N. (2003) Control of sugarcane smut disease in Nigeria with fungicides. *Crop Prot.* **22**, 45–49.

Albert, H.H. and Schenck, S. (1996) Amplification from a homolog of the bE mating-type gene as a sensitive assay for the presence of *Ustilago scitaminea* DNA. *Plant Dis.* **80**, 1198–1192.

Amalraj, A. V and Balasundaram, N. (2006) On the taxonomy of the members of "Saccharum Complex." *Genet. Resour. Crop Evol.* **53**, 35–41.

Antoine, R. (1961) Smut. In Sugarcane diseases of the world. (Martin JP, Abbott EV, H.C., ed), pp. 327–354. Amsterdam: Elsevier.

Bakkeren, G., Kämper, J. and Schirawski, J. (2008) Sex in smut fungi: Structure, function and evolution of mating-type complexes.

Barnabas, L., Ashwin, N.M.R., Kaverinathan, K., et al. (2016) Proteomic analysis of a compatible interaction between sugarcane and *Sporisorium scitamineum*. *Proteomics* **16**, 1111–22.

Bhuiyan, S.A., Croft, B.J., James, R.S. and Cox, M.C. (2012) Laboratory and field evaluation of fungicides for the management of sugarcane smut caused by *Sporisorium scitamineum* in seedcane. *Australas. Plant Pathol.* **41**, 591–599.

Bhuiyan, S.A., Croft, B.J. and Tucker, G.R. (2015) New method of controlling sugarcane smut using Flutriafol fungicide. *Plant Dis.* **99**, 1367–1373.

Borrás-Hidalgo, O., Thomma, B.P.H.J., Carmona, E., Borroto, C.J., Pujol, M., Arencibia, A. and Lopez, J. (2005) Identification of sugarcane genes induced in disease-resistant somaclones upon inoculation with *Ustilago scitaminea* or *Bipolaris sacchari*. *Plant Physiol. Biochem.* **43**, 1115–1121.

Botha, F.C. and Moore, P.H. (2014) Biomass and bioenergy. In Sugarcane - Physiology, biochemistry and functional biology. (Moore, P.H. and Botha, F.C., eds), pp. 521–541. Oxford, UK: Wiley-Blackwell.

Bremer, G. (1961) Problems in breeding and cytology of sugar cane. *Euphytica* **10**, 59–78.

Bueno, C.R.N.C. (2010) *Infection by Sporisorium scitamineum on sugarcane: influence of environmental variables and development of a method for early diagnosis*. Universidade de São Paulo.

Cardoso-Silva, C.B., Costa, E.A., Mancini, M.C., et al. (2014) De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS One* **9**, e88462.

Carvalho, G., Quecine, M.C., Longatto, D.P., et al. (2016) *Sporisorium scitamineum* colonisation of sugarcane genotypes susceptible and resistant to smut revealed by GFP-tagged strains. *Ann. Appl. Biol.* **169**, 329–341.

Cheavegatti-Gianotto, A., Abreu, H.M.C. de, Arruda, P., et al. (2011) Sugarcane (Saccharum X officinarum): a reference study for the regulation of genetically modified cultivars in Brazil. *Trop. Plant Biol.* **4**, 62–89.

Croft, B.J., Berding, N., Cox, M.C. and Bhuiyan, S. (2008) Breeding smut-resistant sugarcane varieties in Australia: progress and future directions. **30**, 125–134.

D'Hont, A. (2005) Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet. Genome Res.* **109**, 27–33.

D'Hont, A. and Glaszmann, J.C. (2001) Sugarcane genome analysis with molecular markers: a first decade of

research. *Proc Int Soc Sugar Cane Technol* **24**, 556–559.

D'Hont, A., Grivet, L., Feldmann, P., Glaszmann, J.C., Rao, S. and Berding, N. (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol. Gen. Genet. MGG* **250**, 405–413.

Daniels, J. and Roach, B.T. (1987) Taxonomy and Evolution. In Sugarcane improvement through breeding. (Heinz, D., ed), pp. 7–84. Amsterdam: Elsevier.

Elston, D.A. and Simmonds, N.W. (1988) Models of sugarcane smut disease and their implications for testing variety resistance. *J. Appl. Ecol.* **25**, 319–329.

Esh, A. (2014) The activity of pathogenesis related proteins in smut resistant and susceptible sugarcane (GT54-9) mutants induced by gamma radiation. *Adv. Plants Agric. Res.* **1**, 1–12.

FAO (2013) Food and Agriculture Organization of United Nations - FAOSTAT. *Production.*

Fontaniella, B., Márquez, A., Rodríguez, C.W., Piñón, D., Solas, M.T., Vicente, C. and Legaz, M.E. (2002) A role for sugarcane glycoproteins in the resistance of sugarcane to *Ustilago scitaminea*. *Plant Physiol. Biochem.* **40**, 881–889.

Goldemberg, J. (2008) The Brazilian biofuels industry. *Biotechnol. Biofuels* **1**, 1–7.

Gupta, S.C. (1978) Control of diseases by hot water treatment of sugarcane seed material. *Indian Sugar Crop. J.* **5**, 28–29.

Hector, E., Prada, F. de and Rodriguez, R. (1995) Experimental evidence for the presence of different smut resistance mechanisms in sugarcane. In Proceedings XXI Congress of ISSCT., p. 15.

Heinze, B., Thokoane, L., Williams, N., Barnes, J. and Rutherford, R. (2001) The smut-sugarcane interaction as a model system for the integration of marker discovery and gene isolation. *Proc. South African Sugar Technol. Assoc.* **75**, 104–107.

Huang, N., Zhang, Y.Y., Xiao, X.H., Huang, L., Wu, Q.B., Que, Y.X. and Xu, L.P. (2015) Identification of smut-responsive genes in sugarcane using cDNA-SRAP. *Genet Mol Res* **14**, 6808–6818.

Jose, R.C., Louis, B., Goyari, S., Waikhom, S.D., Handique, P.J. and Talukdar, N.C. (2016) Biotrophic interaction of *Sporisorium scitamineum* on a new host—*Saccharum spontaneum*. *Micron* **81**, 8–15.

Kavitha, M., Sundar, A.R., Padmanaban, P., Viswanathan, R. and Malathi, P. (2014) Comparative study on early detection of sugarcane smut (*Sporisorium scitamineum*) by polymerase chain reaction (PCR) and microscopy. *African J. Biotechnol.* **13**, 4635–4638.

LaO, M., Arencibia, A., Carmona, E., Acevedo, R., Rodríguez, E., León, O. and Santana, I. (2008) Differential expression analysis by cDNA-AFLP of Saccharum spp. after inoculation with the host pathogen *Sporisorium scitamineum*. *Plant Cell Rep.* **27**, 1103–1111.

Latiza, A.S., Ampusta, D.C. and Rivera, J.R. (1980) Reaction of sugarcane clones to strain B of *Sporisorium scitamineum*. *Int Soc Sugarcane Technol.* **2**, 1456–1462.

Legaz, M.E., Armas, R. d., Pinon, D. and Vicente, C. (1998) Relationships between phenolics-conjugated polyamines and sensitivity of sugarcane to smut (*Ustilago scitaminea*). *J. Exp. Bot.* **49**, 1723–1728.

Lemma, A., Hadush, H., Yohannes, Z. and Amrote, T. (2015) Study on the reaction of sugarcane genotypes (CIRAD-2011) to sugarcane smut (*Sporisorium scitamineum*) in the Ethiopian sugarcane plantations. *Adv. Crop Sci. Technol.* **3**, 1–3.

Lloyd, H.L. and Naidoo, G. (1983) Chemical-assay potentially suitable for determination of smut resistance of sugarcane cultivars. *Plant Dis.* **67**, 1103–1105.

Lloyd, H.L. and Pillay, M. (1980) The development of an improved method for evaluating sugarcane for resistance to smut. *Proc. South African Sugar Technol. Assoc.* **54**, 168–172.

Magarey, B.R., Sheahan, T., Sefton, M., Kerkwyk, R., Bull, J. and Limited, B. (2011) Yield losses from sugarcane smut: indications from Herbert Mill data and from individual crop assessments. *Proc Aust Soc Sugar Cane Technol* **33**, 1–9.

Marques, J.P.R., Appezzato-da-Glória, B., Piepenbring, M., Massola, N.S., Monteiro-Vitorello, C.B. and Vieira, M.L.C. (2016) Sugarcane smut: shedding light on the development of the whip-shaped sorus. *Ann. Bot.*, 1–13.

McNeil, M.D., Hermann, S., Jackson, P.A. and Aitken, K.S. (2011) Conversion of AFLP markers to high-throughput markers in a complex polyploid, sugarcane. *Mol. Breed.* **27**, 395–407.

Millanes, A.M., Fontaniella, B., Legaz, M.E. and Vicente, C. (2005) Glycoproteins from sugarcane plants regulate cell polarity of *Ustilago scitaminea* teliospores. *J. Plant Physiol.* **162**, 253–265.

Nalawade, S.V., Indi, D.V., Pandhare, R.A. and Pawar, S.M. (2012) Evaluation of sugarcane genotypes for resistance to whip smut caused by *Ustilago scitaminea*. *Ecol. Environ. Conserv.* **18**, 551–553.

Olufolaji, D.B. (1993) Evaluation of some relatively new fungicides for smut control in sugar-cane. *Crop Prot.* **12**, 293–295.

Piñon, D., Armas, R. de, Vicente, C. and Legaz, M.E. (1999) Role of polyamines in the infection of sugarcane buds by *Ustilago scitaminea* spores. *Plant Physiol. Biochem.* **37**, 57–64.

Piperidis, G., Piperidis, N. and D'Hont, A. (2010) Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol. Genet. Genomics* **284**, 65–73.

Que, Y., Su, Y., Guo, J., Wu, Q. and Xu, L. (2014) A global view of transcriptome dynamics during *Sporisorium scitamineum* challenge in sugarcane by RNA-seq Fu, B., ed. *PLoS One* **9**, e106476.

Que, Y., Xu, L., Lin, J., Ruan, M., Zhang, M. and Chen, R. (2011) Differential protein expression in sugarcane during sugarcane- *Sporisorium scitamineum* interaction revealed by 2-DE and MALDI-TOF-TOF/MS. *Comp. Funct. Genomics* **2011**.

Roach, B.T. (1972) Nobilization of sugarcane. *Proc. Int. Soc. Sugar Cane Technol* **14**, 206–216.

Rott, P. (2000) A guide to sugarcane diseases, Cirad.

Santiago, R., Alarcón, B., Armas, R. de, Vicente, C. and Legaz, M.E. (2012) Changes in cinnamyl alcohol dehydrogenase activities from sugarcane cultivars inoculated with *Sporisorium scitamineum* sporidia. *Physiol. Plant.* **145**, 245–259.

Santiago, R., Armas, R. de, Blanch, M., Vicente, C. and Legaz, M.-E. (2010) In vitro effects of caffeic acid upon growth of the fungi *Sporisorium scitamineum*. *J. Plant Interact.* **5**, 233–240.

Santiago, R., Armas, R. de, Fontaniella, B., Vicente, C. and Legaz, M.-E. (2009) Changes in soluble and cell wall-bound hydroxycinnamic and hydroxybenzoic acids in sugarcane cultivars inoculated with *Sporisorium scitamineum* sporidia. *Eur. J. Plant Pathol.* **124**, 439–450.

Schenck, S. (2003) New race of sugarcane smut on Maui. *Hawaii Agric. Res. Cent.* **69**, 1–4.

Scortecci, K., Creste, S., Calsa, T., Xavier, M.A., Landell, M., Figueira, A. and Benedito, V.A. (2012) Challenges, Opportunities and Recent Advances in Sugarcane Breeding. In Plant Breeding. (Abdurakhmonov, I.Y., ed), pp. 267–296. Rijeka: InTech.

Setta, N. de, Monteiro-Vitorello, C., Metcalfe, C., et al. (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* **15**, 540.

Singh, N., Somai, B.M. and Pillay, D. (2005) In vitro screening of sugarcane to evaluate smut susceptibility. *Plant Cell. Tissue Organ Cult.* **80**, 259–266.

Singh, N., Somai, B.M. and Pillay, D. (2004) Smut disease assessment by PCR and microscopy in inoculated tissue cultured sugarcane cultivars. *Plant Sci.* **167**, 987–994.

Sinha, O.K., Singh, K. and Misra, S.R. (1982) Stain technique for detection of smut hyphae in nodal buds of sugarcane. *Plant Dis.* **66**, 932–933.

Sreenivasan, T.A.B.H.D. (1987) Cytogenetics. In Sugarcane improvement through breeding. (Heinz, D., ed), pp. 211–254. Amsterdam: Elsevier.

Stoll, M., Begerow, D. and Oberwinkler, F. (2005) Molecular phylogeny of *Ustilago, Sporisorium,* and related taxa based on combined analyses of rDNA sequences. *Mycol. Res.* **109**, 342–56.

Su, Y., Wang, S., Guo, J., Xue, B., Xu, L. and Que, Y. (2013) A TaqMan real-time PCR assay for detection and quantification of *Sporisorium scitamineum* in sugarcane. *Sci. World J.* **2013**, 1–9.

Su, Y., Xu, L., Wang, Z., Peng, Q., Yang, Y., Chen, Y. and Que, Y. (2016) Comparative proteomics reveals that central metabolism changes are associated with resistance against *Sporisorium scitamineum* in sugarcane. *BMC Genomics* **17**, 800.

Su, Y., Xu, L., Xue, B., Wu, Q., Guo, J., Wu, L. and Que, Y. (2013) Molecular cloning and characterization of two pathogenesis-related β-1,3-glucanase genes ScGluA1 and ScGluD1 from sugarcane infected by *Sporisorium scitamineum*. *Plant Cell Rep.* **32**, 1503–1519.

Sundar, A.R., Ashwin, N.M.R., Barnabas, E.L., Malathi, P. and Viswanathan, R. (2015) Disease resistance in sugarcane - an overview. *Sci. Agrar. Parana.* **14**, 200–212.

Sundar, A.R., Barnabas, E.L., Malathi, P. and Viswanathan, R. (2012) A mini-review on smut disease of sugarcane caused by *Sporisorium scitamineum*. In Botany. (Mworia, J.K., ed), pp. 107–128. InTech.

Taniguti, L.M., Schaker, P.D.C., Benevenuto, J., et al. (2015) Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with Sugarcane. *PLoS One* **10**, e0129318.

Thokoane, L.N. and Rutherford, R.S. (2001) cDNA-AFLP differential display of sugarcane (*Saccharum* spp. hybrids) genes induced by challenge with the fungal pathogen *Ustilago scitaminea (*sugarcane smut). *Proc. South African Sugar Technol. Assoc.* **75**, 104–107.

UNICA (2016) União Nacional da Industria da Cana-de-açúcar.

Vettore, A.L., Silva, F.R. da, Kemper, E.L., et al. (2003) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res.* **13**, 2725–2735.

Vettore, A.L., Silva, F.R. da, Kemper, E.L. and Arruda, P. (2001) The libraries that made SUCEST. *Genet. Mol. Biol.* **24**, 1–7.

Waclawovsky, A.J., Sato, P.M., Lembke, C.G., Moore, P.H. and Souza, G.M. (2010) Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnol. J.* **8**, 263–76.

Wada, A.C., Anaso, A.B. and Bassey, M.S. (2016) Sugar cane whip smut (*Sporisorium scitamineum* Syd) caused field sucrose and juice quality losses of two sugar cane varieties in Nigeria. *Int. J. Plant Soil Sci.* **10**, 1–11.

Waller, J.M. (1970) Sugarcane smut (*Ustilago scitaminea*) in Kenya. *Trans. Br. Mycol. Soc.* **54**, 405–414.

Wu, Q., Xu, L., Guo, J., Su, Y. and Que, Y. (2013) Transcriptome profile analysis of sugarcane responses to *Sporisorium scitaminea* infection using Solexa sequencing technology. *Biomed Res. Int.* **2013**, 1–9.

Yan, M., Zhu, G., Lin, S., et al. (2016) The mating-type locus b of the sugarcane smut *Sporisorium scitamineum* is essential for mating, filamentous growth and pathogenicity. *Fungal Genet. Biol.* **86**, 1–8.

You-Xiong, Q., Jian-Wei, L., Xian-Xian, S., Li-Ping, X. and Ru-Kai, C. (2011) Differential gene expression in sugarcane in response to challenge by fungal pathogen *Ustilago scitaminea* revealed by cDNA-AFLP. *J. Biomed. Biotechnol.* **2011**, 1–10.

2. MODULATION OF *SPORISORIUM SCITAMINEUM* GENE EXPRESSION PROFILE RELATED TO SUGARCANE COLONIZATION

Abstract

As all living organisms, plant pathogens modulate gene expression profiles in response to stimuli. This is especially important during the interaction with hosts, since it determines the success of infection. Here, we used RNAseq technique to analyze *Sporisorium scitamineum* gene expression in contrasting growth conditions. The data revealed putative mechanisms activated by the pathogen that allows entrance and survival inside sugarcane tissues, as well as genes that encode proteins that may act as effectors. The results presented in this chapter were published as part of the article entitled "Complete Genome Sequence of *Sporisorium scitamineum* and Biotrophic Interaction Transcriptome with Sugarcane" in PLoS ONE journal, DOI:10.1371/journal.pone.0129318 in June 12, 2015. No permission is required from the authors or the publishers to reuse or repurpose PLOS content provided the original article.

2.1. Introduction

The diseases known as "smut" are caused by fungi of the family Ustilaginaceae (Phylum Basdiomycota, Class Ustilaginomycetes, Order Ustilaginales), comprising about 1200 species (Agrios, 2005; Bakkeren *et al.*, 2008). So far it has been described about 4000 species of host plants (Bakkeren *et al.*, 2008; Martínez-Espinoza *et al.*, 2002), including the most important crops in the world, such as corn, barley, wheat, oats, sorghum, forage grasses and sugarcane (Bakkeren *et al.*, 2008). These fungi develop mainly in the inflorescence, leaves and stems (Stoll *et al.*, 2005), giving rise to fruiting structures containing black mass of teliospores that give the infected tissue an aspect of "soot", so called "smut" (Bakkeren *et al.*, 2008). Most smut fungus infects only a few

susceptible hosts and phylogenetically near, or are restricted to just a single plant species (Begerow *et al.*, 2006).

Sporisorium scitamineum (Syd.) M. Piepenbr., M. Stoll & Oberw is the causal agent of sugarcane smut, one of the major diseases that affect the culture in the world (Sundar et al., 2015), featuring a biotrophic lifestyle with inter- or intracellular growth (Stoll *et al.*, 2005). GFP-fusioned mutants of *S. scitamineum* infecting tissues of sugarcane genotypes with distinct response to smut evidenced its growth even in the resistant varieties and disrupted organization of vascular vessels (Carvalho *et al.*, 2016).

Recently, *S. scitamineum* genome was sequenced by two independent groups. The first release was in 2014, where the genome sequence was *de novo* assembled in 321 contigs and 58 scaffolds, generating an estimated number of 6,636 genes and 68 candidates for secreted effector proteins (Que *et al.*, 2014). Later, a Brazilian strain was sequenced from telomere to telomere using a combination of Illumina and PacBio sequencing (Taniguti *et al.*, 2015). Were assembled 26 chromosomes, 23 of them confirmed by hybridization using telomere probes and by the presence of sequence telomere motifs. The genome was estimated to have 6,677 protein coding genes, with GO terms assigned to 3,682 (55.2%) of them, and top hit species (5,078) with *S. reillianum* proteins. Additionally, *S. scitamineum* genome encodes 527 predicted proteins showing signal peptides, of these 342 have no transmembrane domains and 305 are also not anchored by GPI (Glycosylphosphatidylinositol). This set of genes were assigned as the *S. scitamineum* secretome. Of the secretome proteins, 48.5% have no characterized function and 29 are *S. scitamineum* singletons (Taniguti *et al.*, 2015).

The release of these genomes allows now the discovery of potential genes involved in a successful plant-pathogen interaction. In this sense, determine the genes which present increase expression during the interaction is an interesting approach. Transcriptomics field has been developed rapidly with the advent of next-generation sequencing technologies and generates information about which genes are being expressed, at what level and can also provide information about different transcript isoforms (McGettigan, 2013). In the plant-pathogen interaction, transcriptomics approach was used by several groups to describe pathogen genes activated during interaction (Ailloud *et al.*, 2016; Castell-Miller *et al.*, 2016). For the "smut" pathogens large scale analysis of gene expression was applied through the use of microarrays to identify genes differentially regulated during teliospore germination (Zahiri *et al.*, 2005) and using suppressive subtractive hybridization (SSH) cDNA library from *Ustilago maydis*-infected *Zea mays* to identify the highly expressed genes in planta (Donaldson et al., 2013a) and also to determine expression in mutant lines (Islamovic et al., 2015; Morrison et al., 2012). In this work,

we aimed to describe *S. scitamineum* responsive genes to the plant environment using RNAseq technology and associate to their possible roles as effectors or pathogenicity factors. To our knowledge this was the first work on *S. scitamineum* transcriptomics.

2.2. Material and Methods

2.2.1. Experimental design and RNA extraction

To assess *S. scitamineum* gene expression during its growth in sugarcane and *in vitro* we used RNA-Seq technique (Figure 3). The data of fungal gene expression *in planta* was previously obtained by (Palhares, 2014). Therefore, single budded sets of the susceptible variety "RB925345" were subjected to disinfection by heat and chemical treatment (52°C water bath for 30 min, bath in sodium hypochlorite solution 0.01% for 10 minutes and wash in distilled water) and kept in a moist chamber for 16 hours at 28°C to stimulate bud sprouting. Prior to inoculation, buds were punctured with a sterile needle aiming to break pre-formed resistance. Teliospores of *S. scitamineum* SSC39 were used as inoculum. As a first step, they were tested for viability by inoculation in YM-agar medium (3 $g.L^{-1}$ yeast extract, 3 $g.L^{-1}$ malt extract, 5 $g.L^{-1}$ peptone, 3 $g.L^{-1}$ dextrose, 20 $g.L^{-1}$ agar). Overnight grow plates were subjected to germination count of 100 aleatory teliospores in four plates to achieve the percentage of germination.

Inoculation was carried out using a paste of teliospores of *S. scitamineum* SSC39 isolate presenting viability greater than 80%. The setts were placed in trays containing moist vermiculite and kept in a greenhouse with daily irrigation. After five days, they were transferred to vessels containing equal proportions of topsoil and Tropstrato substrate in a randomized design. Sampling was made at 5 DAI (Days After Infection) and 200 DAI (after issuing the whip). The 5 DAI sample was composed of 3 pools 10 buds, while 200 DAI samples were composed of 1 plant per replicate, with sampling at the whip base region. Three replicates were systematically used. RNA extraction of 5 DAI samples was performed using lithium chloride method and 200 DAI samples with TRIzol (Invitrogen) (Palhares, 2014).

S. scitamineum expression was also obtained for its *in vitro* growth. For this purpose, opposite mating types yeasts of *S. scitamineum* SSC39 isolate were cultured liquid in YM medium for 15 hours at 28°C under agitation (200 rpm) in three biological triplicates. For RNA extraction, cells of opposite mating types the were pooled (maintaining the three replicates) and concentrated by centrifugation. RNA extraction was performed using TRIzol reagent according to the manufacturer's instructions. RNA integrity was confirmed in agarose-SYBR safe gels and

concentration and quality ($A_{260/280}$) was estimated by spectrometry (NanoDrop 2009, Thermo Scientific Fisher Inc).

2.2.2. Data collection and analysis

RNAseq libraries were built according to the "TruSeq RNA Sample Prep v2 Low Throughput (LT)" (Illumina) kit instructions and sequenced in HiScanSQ platform (Illumina). The pair-end reads of ~ 100 bp were analyzed by FastqC v. 0.10.1 program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the bases with low quality (Phred <20), adapter sequences, poly-A tails, reads < 50 bp, were eliminated using SeqyClean program (ver. 1.8.10) (http://sourceforge.net/projects/seqclean/files/). Ribosomal RNA sequences were filtered using mappings to the *S. scitamineum* (NCBI access JN367321 and AY550243) and sugarcane (BAC clone) rDNAs using Bowtie2 software (Langmead and Salzberg, 2012).

Trimmed reads from plant inoculated samples were then mapped *S. scitamineum* SSC39 complete genome (Taniguti *et al.*, 2015) using Bowtie2 (Langmead and Salzberg, 2012) to create a set of reads from pathogen, which were used for expression analysis. For this purpose, mappings of pathogen reads were made in CLC Genomics Workbench program (100% identity, 98% coverage) using as reference the 6677 CDSs identified in the genome of the strain *S. scitamineum* SSC39. After, the number of mapped reads was normalized by scaling and CDS length. The mappings results were first used to detected pathogen genes that are most expressed in each of the evaluated conditions, which were called genes preferentially expressed.

As second approach, we determined differentially expressed genes *in planta* compared to *in vitro* growth. Were considered differentially expressed those genes with FDR ≤ 0.01 obtained in Baggerley's test and Log2 Fold Change (*in vitro*/*in planta*) ≤ -2 or ≥ 2. The enrichment of GO terms in the sets of differentially expressed genes was performed in BLAST2GO (Fisher's test, p-value ≤ 0.05).

Finally, data mapping was used to establish genes that were expressed only during the interaction with sugarcane. Interaction specific genes were considered those with no reads mapped into any of the three replicates *in vitro*, and at least one pair of reads mapped in three replicates in the plant growth.

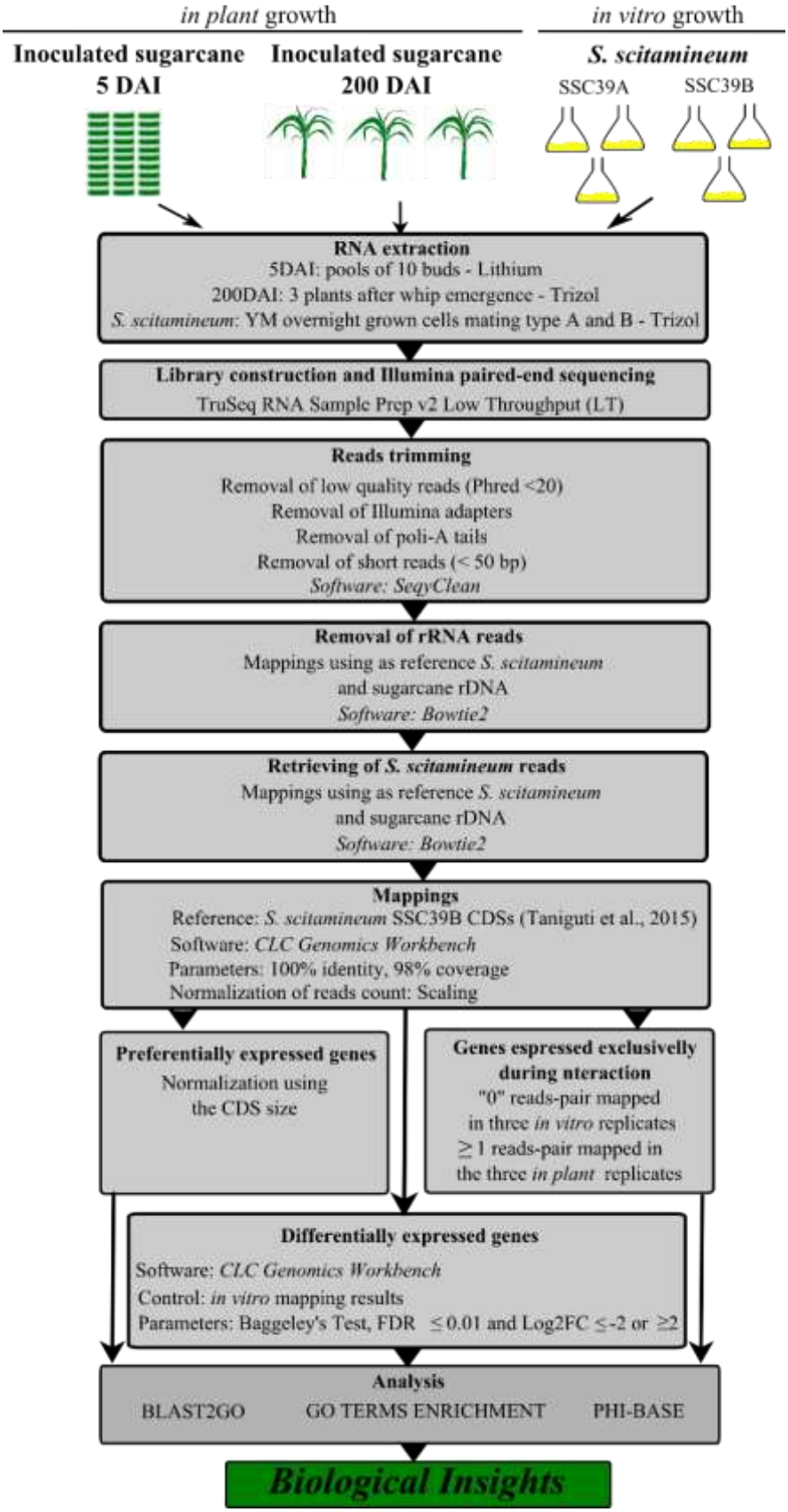


Figure 3. Schematic representation of RNAseq analysis to analyze *S. scitamineum* genes expression *in vitro* and during its interaction with sugarcane.

2.3. Results

To identify fungal genes responsive to the interaction with sugarcane, transcriptional profiles were obtained in three conditions: *in vitro* growth and *in planta* growth of samples 5 and 200 DAI. The data were further used to determine 1) the genes most expressed in each condition, 2) genes differentially expressed during host interaction at 5 DAI and 200 DAI using RNAseq data of *S. scitamineum* growing *in vitro* as control, and 3) genes expressed only during interaction.

Table 1 shows the total number of reads obtained and the remaining *S. scitamineum* reads after trimming. Approximately 56 %, 39 % and 53 % of reads mapped to *S. scitamineum* CDSs *in vitro*, 5 DAI and 200 DAI, respectively. The number of CDSs transcribed in each experiment were 6,213 CDSs (93%) *in vitro*, 437 CDSs (6.5%) at 5 DAI and 6,183 CDSs (92.6%) at 200 DAI. Of the non-transcribed genes *in vitro* (464) and *in planta* (493), 333 are shared by the two sets, representing 4.9% of the total number of predicted genes of *S. scitamineum* genome. In this set of genes none GO term is enriched (p-value < 0.05) and annotation process (Figure 4A) indicates that few ones have predicted functions related mainly to carbohydrate and nitrogen processes (Figure 4B). They may be expressed in conditions other than the used in our experiments, erroneous gene predictions or low sequencing coverage of the experiments *in planta*.

Table 1. Results of transcripts sequencing, trimming and mappings using *S. scitamineum* CDS as reference.

Condition	Total number of paired-end reads	Total number of paired-end reads after trimming	Mapped reads in pairs
Culture medium YM _1	39,383,080	15,073,414	8,377,340
Culture medium YM _2	47,140,252	17,703,654	10,000,564
Culture medium YM _3	41,945,644	14,755,906	8,542,818
Sugarcane 5 DAI_1	16,266,804	12,416	4,824
Sugarcane 5 DAI_2	21,017,222	23,194	10,634
Sugarcane 5 DAI_3	19,075,478	11,468	4,016
Sugarcane 200 DAI_1	17,705,938	1,088,146	558,322
Sugarcane 200 DAI_2	21,691,820	3,605,114	1,990,300
Sugarcane 200 DAI_3	17,617,212	2,088,044	1,164,430

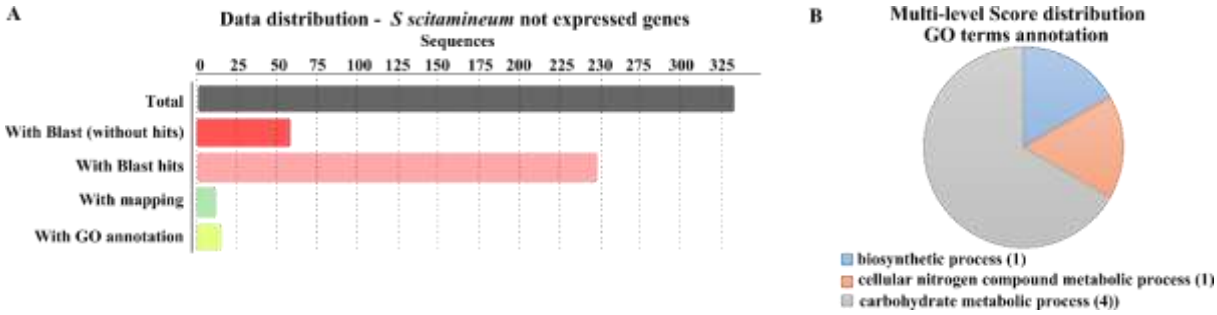


Figure 4. Blast2GO analysis of *S. scitamineum* genes not expressed in any of the tested conditions: *in vitro* growth – YM media, and *in plant* growth – early infection and after whip development. A) Data distribution and B) multi-level distribution of assigned GO-terms.

2.3.1. Genes preferentially expressed

Considering the relatively low number of fungal reads recovered of the experiment *in planta* 5 DAI, we analyzed the most expressed genes in each treatment according to the number of mapping reads per CDS normalized by scaling approach and gene length (Kbp). In this case, again only genes mapped by at least one pair of reads in all three replicates were considered. We called these genes preferentially expressed (Figure 5, Table 2). The mapping of reads to genes preferentially expressed were in some cases more than 180 times higher than the average number of reads mapped per CDS.

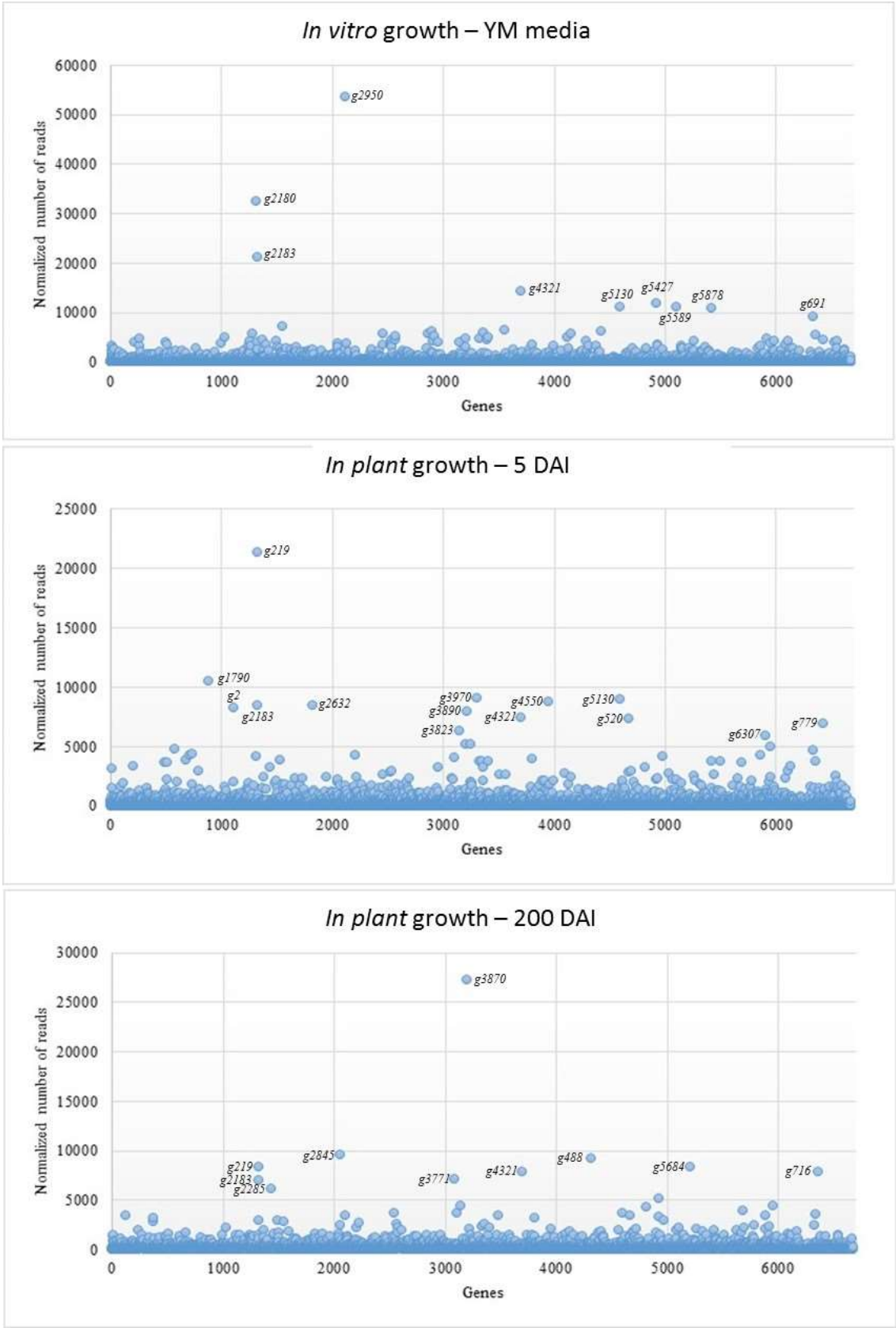


Figure 5. Graphic view of *S. scitamineum* expression in each growth condition analyzed. Each dot represents one gene.

Table 2. *S. scitamineum* fifth most expressed genes in each condition analyzed. "aa": number of amino acids in the mature codified protein. "PS": presence (Y) or not (N) of peptide signal.

	YM – culture medium				Sugarcane - 5 DAI				Sugarcane - 200 DAI			
	Gene	**Annotation**	**aa**	**PS**	**Gene**	**Annotation**	**aa**	**PS**	**Gene**	**Annotation**	**aa**	**PS**
1	g2905	Alternative oxidase. mitochondrial	402	N	g219	12 kDa heat shock protein	79	N	g3870	Uncharacterized protein	283	Y
2	g2180	ADP.ATP carrier protein	317	N	g1790	Alcohol dehydrogenase	370	N	g2845	Histone H4	103	N
3	g2183	Elongation factor 1-alpha	459	N	g3970	Uncharacterized protein	745	Y	g488	Uncharacterized protein	294	Y
4	g4321	Polyubiquitin	229	N	g5130	Peptidyl-prolyl cis-trans isomerase	179	N	g219	12 kDa heat shock protein	79	N
5	g5427	Uncharacterized protein	201	N	g4550	Uncharacterized protein	179	N	g5684	Uncharacterized protein	531	Y
6	g5130	Peptidyl-prolyl cis-trans isomerase	179	N	g2632	Aldehyde dehydrogenase	497	N	g4321	Polyubiquitin	229	N
7	g5589	Ubiquitin-40S ribosomal protein	159	N	g2183	Elongation factor 1-alpha	459	N	g716	Probable thiamine biosynthesis protein	359	N
8	g5878	Uncharacterized oxidoreductase	324	N	g2	Uncharacterized protein	236	Y	g3771	Uncharacterized protein	189	N
9	g691	Glyceraldehyde-3-phosphate dehydrogenase	337	N	g3890	Uncharacterized protein	135	Y	g2183	Elongation factor 1-alpha	459	N
10	g2391	40S ribosomal protein	145	N	g520	Uncharacterized protein	97	N	g2285	Uncharacterized protein	243	N
11	g419	Malate dehydrogenase. mitochondrial	340	N	g4321	Polyubiquitin	229	N	g5434	Histone H3	136	N
12	g3602	Phosphoenolpyruvate carboxykinase [ATP]	518	N	g779	Uncharacterized protein	151	N	g3823	Probable heat shock protein 80	678	N
13	g4982	ATP synthase subunit alpha. mitochondrial	543	N	g3823	Probable heat shock protein 80	678	N	g6354	Probable NADP-dependent mannitol dehydrogenase	261	N
14	g4018	60S ribosomal protein	220	N	g6307	Uncharacterized protein	321	Y	g5332	Uncharacterized protein	317	N
15	g2145	60S ribosomal protein	154	N	g3920	Related to Mig1 protein	202	Y	g6107	Uncharacterized protein	346	Y
16	g4734	Glutamine synthetase	355	N	g388	Opsin-1	292	N	g3790	Endoglucanase	379	Y
17	g3208	Monothiol glutaredoxin-5. mitochondrial	158	N	g6354	Probable NADP-dependent mannitol dehydrogenase	261	N	g5130	Peptidyl-prolyl cis-trans isomerase	179	N
18	g3576	Mitochondrial phosphate carrier protein	325	N	g1513	Uncharacterized protein	215	Y	g3285	Uncharacterized protein	454	N
19	g716	Probable thiamine biosynthesis protein	359	N	g691	Glyceraldehyde-3-phosphate dehydrogenase	337	N	g691	Glyceraldehyde-3-phosphate dehydrogenase	337	N
20	g3312	Translocator protein homolog	190	N	g1655	Probable alpha-L-arabinofuranosidase A	374	N	g2887	Peroxygenase 2	246	N
21	g3626	Uncharacterized protein	255	Y	g627	Uncharacterized protein	291	N	g6293	Uncharacterized protein	110	Y
22	g1925	Actin	375	N	g2980	Superoxide dismutase [Mn]	182	N	g520	Uncharacterized protein	97	N
23	g4058	Enolase	445	N	g1642	Uncharacterized protein	1169	N	g4124	Related to monocarboxylate permease	443	N
24	g4706	Prohibitin-1	268	Y	g5475	30 kDa heat shock protein	207	N	g1102	Thiamine thiazole synthase	323	N
25	g3995	Triosephosphate isomerase	248	N	g2180	ADP.ATP carrier protein	317	N	g5427	Uncharacterized protein	201	N
26	g1234	Heat shock protein 60	552	N	g3790	Endoglucanase 1	379	Y	g4419	Transaldolase	321	N
27	g6326	Mitochondrial outer membrane protein porin	296	N	g4419	Transaldolase	321	N	g1334	Histone H2B	225	N
28	g3878	ATP synthase subunit beta. mitochondrial	506	N	g1612	N-acetylglucosamine deacetylase	279	Y	g2337	Uncharacterized protein	121	Y
29	g4047	Uncharacterized protein	495	N	g2374	Uncharacterized protein	172	N	g5475	30 kDa heat shock protein	207	N
30	g2230	40S ribosomal protein	145	N	g3990	Uncharacterized protein	222	N	g2180	ADP.ATP carrier protein	317	N
31	g3273	40S ribosomal protein	101	N	g716	Probable thiamine biosynthesis protein	359	N	g1333	Histone H2A	136	N
32	g3309	Uncharacterized protein	123	N	g4008	Uncharacterized protein	238	N	g2391	40S ribosomal protein	145	N
33	g779	Uncharacterized protein	151	N	g5878	Related to 2.5-diketo-D-gluconic acid reductase	324	N	g2998	Probable quinone oxidoreductase	358	N
34	g6495	60S acidic ribosomal protein	313	N	g4058	Enolase	445	N	g4018	60S ribosomal protein	220	N
35	g5733	60S ribosomal protein	216	N	g5941	Endo-1.4-beta-xylanase	756	Y	g3309	Uncharacterized protein	123	N

	YM – culture medium				Sugarcane - 5 DAI				Sugarcane - 200 DAI			
	Gene	**Annotation**	**aa**	**PS**	**Gene**	**Annotation**	**aa**	**PS**	**Gene**	**Annotation**	**aa**	**PS**
36	g4858	Pyruvate dehydrogenase E1 component subunit alpha	411	N	g1455	60S ribosomal protein	195	N	g2846	Histone H3	136	N
37	g882	40S ribosomal protein	243	N	g143	Uncharacterized protein	514	Y	g688	Uncharacterized protein	630	N
38	g6386	Uncharacterized protein	244	N	g6118	Oxidoreductase	315	N	g620	Uncharacterized protein	643	N
39	g2345	14-3-3 protein homolog	261	N	g1183	Ammonium transporter	484	N	g6325	Uncharacterized protein	189	N
40	g2127	Uncharacterized protein	291	Y	g6511	Inorganic pyrophosphatase	329	N	g3995	Triosephosphate isomerase	248	N
41	g3613	Putative nucleosome assembly protein	416	N	g4021	Uncharacterized protein	209	N	g5589	Probable ubiquitin/ribosomal protein S27a fusion protein	159	N
42	g1448	Citrate synthase. mitochondrial	474	N	g3659	40S ribosomal protein	169	N	g2980	Superoxide dismutase [Mn]. mitochondrial	182	N
43	g2185	60S ribosomal protein	130	N	g2285	Uncharacterized protein	243	N	g6118	Uncharacterized oxidoreductase	315	N
44	g3659	40S ribosomal protein	169	N	g5332	Uncharacterized protein	317	N	g3312	Related to Peripheral-type benzodiazepine receptor	190	N
45	g3823	Heat shock protein 90-1	678	N	g1004	Protein transport protein	208	N	g4058	Enolase	445	N
46	g1187	Methylsterol monooxygenase	313	N	g6495	60S acidic ribosomal protein	313	N	g1925	Actin	375	N
47	g865	Uricase	408	N	g5222	Uncharacterized protein	174	N	g6298	Uncharacterized protein	770	N
48	g2835	Cytochrome c	176	N	g3282	Uncharacterized protein	402	Y	g5733	60S ribosomal protein	216	N
49	g2907	Nucleoside diphosphate kinase	152	N	g1713	Zinc-type alcohol dehydrogenase-like protein	354	N	g5249	Zinc-type alcohol dehydrogenase-like protein	435	N
50	g1893	Probable Mrb1-Mitochondrial p32 Family Protein	266	N	g3272	ATP synthase subunit d. mitochondrial	170	N	g	Uncharacterized protein	135	Y

Among the fifth preferentially expressed genes in each condition (Table 2), the number of those encoding probable secreted proteins, i.e. are part of pathogen secretome, is higher *in planta* (17) than *in vitro* (3). Among them, only two genes are shared by the *in planta* samples (5 DAI and 200 DAI): the endoglucanase (g3790_chr09_Ss) and an uncharacterized secreted protein (g3890_chr10_Ss), indicating that the pathogen modules the transcripts profile according to disease stage, which may be related to colonization and survive in the early stages of disease and to whip development in the late stages.

At 5 DAI, the two most expressed genes are a heat shock encoding protein (g219_chr01_Ss) and an alcohol dehydrogenase (g1790_chr03_Ss). The third most expressed gene codify to a secreted protein (g3970_chr10_Ss) that presents several repeats, such as the "PQPQDGQ" motif represented seven times close to the N-terminal region and "PYGDKPNGDAENSDS" repeated eight times towards the C-terminal region (Figure 6). Other four genes (g2_chr01_Ss, g3890_chr10, g6307_chr21_Ss and g1513_chr03_Ss) encode small secreted proteins of 236, 135, 321 and 215 amino acids respectively with no identifiable conserved domains or any sequence feature (Table 2). Additionally, g3890_chr10_Ss expression was detected only *in planta*. The same occurs for g3970_chr10_Ss, an ortholog of Mig1 effectors from *Ustilago maydis,* which is among the most expressed genes at 5 DAI and is only detected *in planta* samples. Genes coding for secreted enzymes endo-1,4-beta-xylanase (g5941_chr19_Ss) and deacetylase (g1612_chr03_Ss) are also highly represented in the RNAseq data 5 DAI (Table 2).

Figure 6. Repeated motifs in g3970_chr_10 gene from *S. scitamineum*, onde of the most expressed during its interaction with sugarcane.

At 200 DAI, some genes highly expressed encode secreted proteins of unknown function that are rich in glycine residues, such as g3870_chr10_Ss and g488_chr01_Ss, which present 26 and 20 % of glycine respectively, derived mainly from "GS", "GKG" and "GEE" repeats in g3870_chr10_Ss and "GEEKK" and "GGE" in g488_chr01_Ss (Figure 7).

Among the other most expressed genes at 200 DAI, g3771_chr09_Ss, g2337_chr05_Ss and g3890_chr10_Ss had not expression detected *in vitro* samples. Additionally, the same superoxide dismutase detected as highly expressed at 5 DAI (g2980_chr07_Ss) is among preferentially expressed genes 200 DAI.

Figure 7. Repeated motifs in g3870_chr10_Ss and g488_chr01_Ss *S. scitamineum* genes that contributes to elevated percentage of glycine residues.

Most of the genes considered preferentially expressed *in vitro* encode proteins related to energetic metabolism and growth, including an alternative oxidase (g2905_chr06_Ss), which is the most expressed, ATP-ADP carrier protein, elongation factor 1-alfa, polyubiquitin and several ribosomal proteins. Seven genes of unknown function are also identified, two of them encoding secreted proteins (g3626_chr09_Ss and g2127_chr04_Ss).

2.3.2. Differentially expressed genes

The results of the differentially expressed genes (DEGs), obtained by CLC comparing each treatment *in planta* with the control *in vitro,* resulted in a total of 125 genes detected as DEGs at 5 DAI, of these 119 are up-regulated and 6 down-regulated *in planta*. At 200 DAI 907 genes were detected as differentially expressed, of these 641 are up-regulated and 266 down-regulated (Figure 8). GO terms assigned to down-regulated genes at 5 DAI are enriched into mRNA binding (GO:0003729) functional group, and processes related to carbohydrate metabolism (GO:0005975), oxidation-reduction (GO:0045333, GO:0015980) and cellular respiration (GO:0045333). Up-regulated genes at 5 DAI are enriched in terms related to transporter activity (GO:0022857, GO:0005215, GO:0055085, GO:0044765, GO:0006810, GO:0006811) and molecular/signal transduction (GO:0060089, GO:0004871). At 200 DAI, the down-regulated genes are enriched in 39 GO terms, including 21 biological process, 12 cellular components and

six molecular functions. The most enriched GO term for the down-regulated genes is catalytic activity (GO:0003824). Up-regulated genes at 200 DAI are enriched in 23 GO terms, four into molecular functions, 14 biological process and five cellular components. The hydrolase activity acting on glycosyl bonds (GO:0016798) and carbohydrate metabolic process (GO:0005975) are the most significant terms enriched (Table 3).

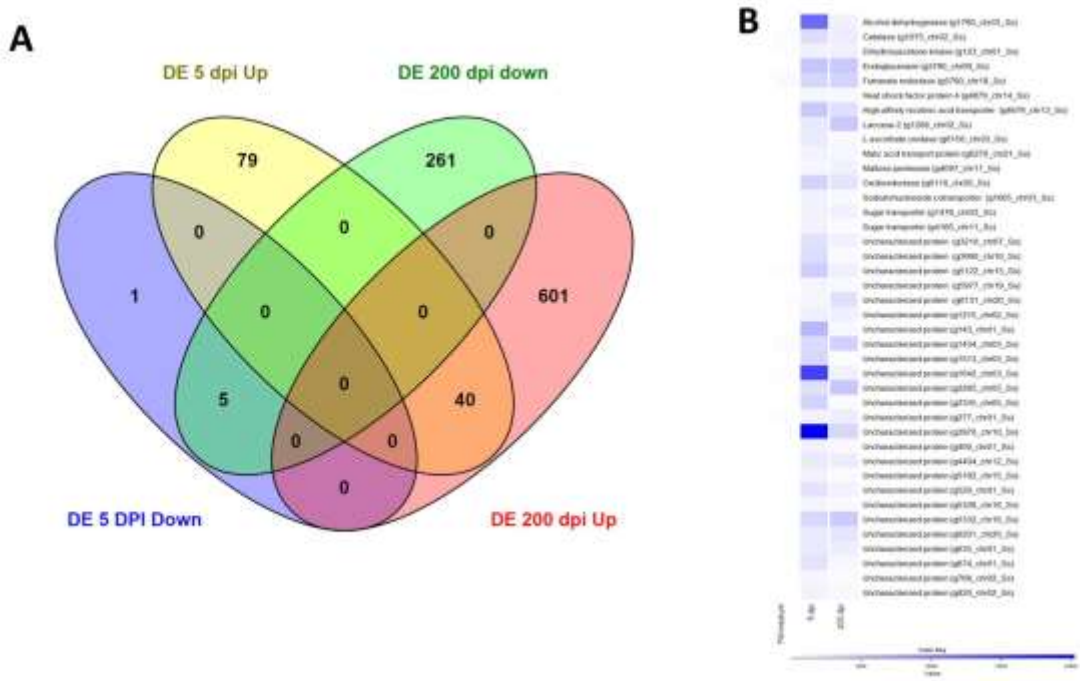


Figure 8. Differential expression analysis of *S. scitamineum* genes. A) Venn diagram of *S. scitamineum* SSC39 differentially expressed during its interaction with sugarcane. B) Heat-map of fungal up-regulated genes in both 5 and 200 DAI.

Table 3. GO terms enrichment in the sets of *S. scitamineum* genes differentially expressed during its grown in sugarcane tissues. F: molecular function, P: biological process, C: cellular components.

	GO-ID	Term	Category	p-value
5 DAI down-regulated	GO:0003729	mRNA binding	F	5.42E-03
	GO:0044444	cytoplasmic part	C	8.69E-03
	GO:0005739	mitochondrion	C	1.09E-02
	GO:0043229	intracellular organelle	C	1.20E-02
	GO:0043226	organelle	C	1.20E-02
	GO:0005975	carbohydrate metabolic process	P	2.07E-02
	GO:0005622	intracellular	C	2.58E-02
	GO:0044424	intracellular part	C	2.58E-02
	GO:0044464	cell part	C	2.67E-02
	GO:0005623	cell	C	2.67E-02
	GO:0015980	energy derivation by oxidation of organic compounds	P	2.82E-02
	GO:0055114	oxidation-reduction process	P	2.82E-02
	GO:0045333	cellular respiration	P	2.82E-02
	GO:0005737	cytoplasm	C	3.70E-02
	GO:0043231	intracellular membrane-bounded organelle	C	4.75E-02
	GO:0043227	membrane-bounded organelle	C	4.75E-02
5 DAI up-regulated	GO:0022857	transmembrane transporter activity	F	1.07E-04
	GO:0005215	transporter activity	F	2.03E-04
	GO:0055085	transmembrane transport	P	6.19E-04
	GO:0016020	membrane	C	9.86E-04
	GO:0044765	single-organism transport	P	1.23E-03
	GO:0060089	molecular transducer activity	F	1.54E-03
	GO:0004871	signal transducer activity	F	1.54E-03

	GO-ID	Term	Category	p-value
	GO:0051234	establishment of localization	P	2.06E-03
	GO:0051179	localization	P	2.06E-03
	GO:0006810	transport	P	2.06E-03
	GO:0006811	ion transport	P	1.12E-02
	GO:0005576	extracellular region	C	4.89E-02
200 DAI down-regulated	GO:0005739	mitochondrion	C	7.26E-14
	GO:0016491	oxidoreductase activity	F	1.25E-08
	GO:0044281	small molecule metabolic process	P	1.04E-06
	GO:0044444	cytoplasmic part	C	1.80E-06
	GO:0044710	single-organism metabolic process	P	2.90E-06
	GO:0005737	cytoplasm	C	6.24E-06
	GO:0006091	generation of precursor metabolites and energy	P	8.37E-06
	GO:0044429	mitochondrial part	C	8.95E-06
	GO:0031975	envelope	C	8.95E-06
	GO:0031967	organelle envelope	C	8.95E-06
	GO:0005740	mitochondrial envelope	C	8.95E-06
	GO:0044763	single-organism cellular process	P	2.53E-05
	GO:0003824	catalytic activity	F	3.70E-05
	GO:0044699	single-organism process	P	5.09E-05
	GO:0043436	oxoacid metabolic process	P	6.63E-05
	GO:1901564	organonitrogen compound metabolic process	P	6.63E-05
	GO:0019752	carboxylic acid metabolic process	P	6.63E-05
	GO:0006520	cellular amino acid metabolic process	P	6.63E-05
	GO:0006082	organic acid metabolic process	P	6.63E-05
	GO:0016829	lyase activity	F	1.31E-04
	GO:0015980	energy derivation by oxidation of organic compounds	P	3.68E-04
	GO:0055114	oxidation-reduction process	P	3.68E-04
	GO:0045333	cellular respiration	P	3.68E-04
	GO:0016746	transferase activity, transferring acyl groups	F	7.90E-04
	GO:0016020	membrane	C	1.86E-03
	GO:0044446	intracellular organelle part	C	2.00E-03
	GO:0044422	organelle part	C	2.00E-03
	GO:0005215	transporter activity	F	4.17E-03
	GO:0055086	nucleobase-containing small molecule metabolic process	P	4.20E-03
	GO:0022857	transmembrane transporter activity	F	4.60E-03
	GO:0005975	carbohydrate metabolic process	P	4.88E-03
	GO:0051186	cofactor metabolic process	P	7.05E-03
	GO:0008152	metabolic process	P	8.14E-03
	GO:0006811	ion transport	P	2.35E-02
	GO:0055085	transmembrane transport	P	3.28E-02
	GO:0006979	response to oxidative stress	P	4.02E-02
	GO:0042579	microbody	C	4.02E-02
	GO:0005777	peroxisome	C	4.02E-02
	GO:0044237	cellular metabolic process	P	4.84E-02
200 DAI up-regulated	GO:0016798	hydrolase activity, acting on glycosyl bonds	F	6.96E-10
	GO:0005975	carbohydrate metabolic process	P	3.87E-08
	GO:0055085	transmembrane transport	P	1.28E-06
	GO:0044765	single-organism transport	P	3.16E-04
	GO:0005694	chromosome	C	6.18E-04
	GO:0051234	establishment of localization	P	9.86E-04
	GO:0051179	localization	P	9.86E-04
	GO:0006810	transport	P	9.86E-04
	GO:0005618	cell wall	C	5.08E-03
	GO:0030312	external encapsulating structure	C	5.08E-03
	GO:0046942	carboxylic acid transport	P	5.42E-03
	GO:0015849	organic acid transport	P	5.42E-03
	GO:0006865	amino acid transport	P	5.42E-03
	GO:0006820	anion transport	P	5.42E-03
	GO:0015711	organic anion transport	P	5.42E-03
	GO:0022857	transmembrane transporter activity	F	5.86E-03
	GO:0071705	nitrogen compound transport	P	1.03E-02
	GO:0071554	cell wall organization or biogenesis	P	1.09E-02
	GO:0005215	transporter activity	F	1.34E-02
	GO:0008643	carbohydrate transport	P	1.54E-02
	GO:0016020	membrane	C	1.92E-02
	GO:0007049	cell cycle	P	3.25E-02
	GO:0016491	oxidoreductase activity	F	3.99E-02

Among differentially expressed genes up-regulated *in planta* (5 and/or 200 DAI) 78 encode proteins which are also secreted (secretome). They are related to host attack, nutrient acquisition and chitin modification, including lipase (g189_chr01_Ss, g4618_chr13_Ss), exo-b-1,3-glucanase (g252_chr01_Ss), b-glucosidase (g468_chr01_Ss, g5316_chr16_Ss), a-L-arabinofuranosidase (g1656_chr03_Ss, g2264_chr04_Ss), pectin lyase (g3529_chr08_Ss), endoglucanase (g3790_chr09_Ss), α-galactosidase (g4463_chr12_Ss), endo-1,4-β-xylanase (g5941_chr19_Ss), FET5-multicopper oxidase (g1208_chr02_Ss), sugar transporters and deacetylase (g6243_chr21_Ss, g1612_chr03_Ss). Genes encoding secreted proteases were also identified as up-regulated such as aspartate protease (g74_chr01_Ss), aspartic protease (g3568_chr09_Ss) and subtilisin-like serine protease (g3042_chr07_Ss) (Table 4).

Transporters encoding genes were differentially expressed *in planta* in both moments, including siderophore transporters (g3806_chr09_Ss, g2279_chr05_Ss), ammonium and nitrate transporters (g4863_chr14_Ss, g1183_chr02_g6016_chr19_Ss, Ss, g5527_chr17_Ss), amino acids and vitamins transport (g5482_chr16_Ss, g2895_chr06_Ss, g5681_chr17_Ss), ABC transporters (g4388_chr12_Ss, g6414_chr21_Ss) and sugar transporters (g4185_chr11_Ss, g1478_chr03_Ss, g1034_chr02_Ss, g4185_chr11_Ss, g1478_chr03_Ss, g6532_chr22_Ss). Invertase was also an important differentially expressed gene at 200 DAI related to carbon acquisition in host interface. Among the differentially expressed genes *in planta,* some are located in subtelomeric regions, such as aldehyde dehydrogenase (g2632_chr06_Ss) and maltose permease (g4097_chr11_Ss) at 5 DAI, and alpha-1,3-mannosyltransferase (g2254_chr04_Ss), sugar transporter (g6215_chr20_Ss), siderophore iron transporter (g3806_chr09_Ss) and brefeldin A resistance protein (g5260_chr15_Ss) at 200 DAI (Table 4).

An ortholog of the gene encoding for the secreted Mig1 protein (g3919_chr10_Ss), was also identified. Another worthy mention secreted protein up-regulated at 200 DAI in *S. scitamineum* is a chorismate mutase (g6307_chr21_Ss), that can interfere with the salicylate mediated plant defense in *U. maydis* (Djame*i et al*, 2011). The salicylate hydroxylase (g4103_chr11_Ss) is, likewise, up-regulated at 200 DAI, and can be related to attenuating salicylate signalization in host tissues (Rabe *et al.*, 2013). Notable is the presence of genes up-regulated *in planta* related to toxin production and detoxification, such as orthologs of versicolorin b synthase (g3941_chr10_Ss) and benzoate 4-monooxygenase (g4198_chr11_Ss), respectively. Besides that, three polyketide synthases potentially involved in toxin biosynthesis (g3298_chr08_Ss, g3302_chr08_Ss and g5915_chr19_Ss) were also up-regulated. Genes related to signal transduction were also up-regulated, including protein kinases (g2874_chr06_Ss, g1321_chr03_Ss, g2134_chr04_Ss, g2002_chr04_Ss, g722_chr02_Ss), transcriptional initiation

factors (g3652_chr09_Ss, g1400_chr03_Ss, g3766_chr09_Ss), transcriptional regulatory proteins (g1809_chr03_Ss), and G-proteins. In addition, 47 of the differentially expressed genes *in planta* encode to proteins which are secreted and of unknown function (Table 4).

Table 4. List of selected *S. scitamineum* genes differentially expressed up-regulated *in plant*.

Secreted			
Process	**Gene ID**	**Annotation**	**Experiment**
Host attack	g74_chr01_Ss	Related to pepsin (Aspartate protease)	200 DAI
	g189_chr01_Ss	Related to Lipase	200 DAI
	g252_chr01_Ss	Glucan 1,3-beta-glucosidase	200 DAI
	g468_chr01_Ss	Probable beta-glucosidase	200 DAI
	g1208_chr02_Ss	Laccase-2	200 DAI/5 DAI
	g1656_chr03_Ss	Alpha-L-arabinofuranosidase	200 DAI
	g1624_chr03_Ss	Guanyl-specific ribonuclease	200 DAI
	g2264_chr04_Ss	Alpha-L-arabinofuranosidase	200 DAI
	g2858_chr06_Ss	Probable lysozyme	200 DAI
	g3042_chr07_Ss	Related to subtilisin-like serine protease	200 DAI
	g3262_chr08_Ss	Related to aminopeptidase	200 DAI
	g3529_chr08_Ss	Related to Pectin lyase	200 DAI
	g3568_chr09_Ss	Related to secreted aspartic protease	200 DAI
	g3696_chr09_Ss	Endo-1,6-beta-D-glucanase	200 DAI
	g3790_chr09_Ss	Endoglucanase	200 DAI/5 DAI
	g3919_chr10_Ss	Related to Mig1 protein	200 DAI
	g4618_chr13_Ss	Lipase	200 DAI
	g5316_chr16_Ss	Probable beta-glucosidase	200 DAI
	g4719_chr13_Ss	Probable pectinesterase	200 DAI
	g5941_chr19_Ss	Endo-1,4-beta-xylanase	200 DAI
	g6000_chr19_Ss	Glucan 1,3-beta-glucosidase	200 DAI
Nutrient acquisition	g4081_chr10_Ss	Related to 3-phytase	200 DAI
	g5690_chr17_Ss	6-hydroxy-D-nicotine oxidase	5 DAI
Chitin modification	g1612_chr03_Ss	N-acetylglucosamine deacetylase	5 DAI
	g1900_chr04_Ss	Chitinase	200 DAI
	g6059_chr20_Ss	Related to Chitin-binding protein	200 DAI
Detoxification	g6307_chr21_Ss	Chorismate mutase	200 DAI
Not Secreted			
Siderophore transporters	g3806_chr09_Ss	Siderophore iron transporter	200 DAI
Ammonium and nitrate transporters	g4863_chr14_Ss	Nitrate transporter	200 DAI
	g1183_chr02_Ss	High affinity ammonium transporter	5 DAI
	g6016_chr19_Ss	Glutathione transporter	200 DAI
	g5527_chr17_Ss	Ammonium transporter	5 DAI
Amino acids and vitamins transport	g5482_chr16_Ss	Dityrosine transporter	200 DAI
	g2895_chr06_Ss	Probable metal-nicotianamine transporter	5 DAI
	g5681_chr17_Ss	Riboflavin transporter	200 DAI
Sugar transporters	g4185_chr11_Ss	Hexose transporter	200 DAI/5 DAI
	g1478_chr03_Ss	Sugar transporter	200 DAI/5 DAI
	g1034_chr02_Ss	High-affinity glucose transporter	200 DAI
	g4185_chr11_Ss	Hexose transporter	200 DAI/5 DAI
	g1478_chr03_Ss	Sugar transporter	200 DAI/5 DAI
	g6532_chr22_Ss	UDP-galactose transporter	200 DAI
Invertase	g1777_chr03_Ss	Invertase	200 DAI
Detoxification	g4103_chr11_Ss	Salicylate hydroxylase	200 DAI
	g4198_chr11_Ss	Pisatin demethylase	200 DAI
Toxin biosynthesis	g3941_chr10_Ss	Versicolorin B synthase	200 DAI
Signal transduction	g2874_chr06_Ss	Hybrid signal transduction histidine kinase	5 DAI
	g1321_chr03_Ss	Serine/threonine-protein kinase	5 DAI
	g2134_chr04_Ss	Serine/threonine-protein kinase	5 DAI
	g2002_chr04_Ss	Probable serine/threonine-protein kinase	200 DAI
	g3652_chr09_Ss	Transcription initiation factor IIA large subunit	5 DAI
	g1400_chr03_Ss	Transcriptional activator of proteases	200 DAI
	g3766_chr09_Ss	Transcription factor RFX4	200 DAI
	g722_chr02_Ss	Serine/threonine-protein kinase	200 DAI
	g1809_chr03_Ss	Transcriptional regulatory protein	200 DAI

2.3.3. Genes expressed exclusively during interaction

Searching for fungal genes expressed only *in planta* we found one gene particular to the interaction at 5 DAI (g4078_chr10_Ss), nevertheless, so far, it encodes an uncharacterized protein, not secreted and with no conserved domains detected. Genes only expressed 200 DAI are 131*:* six of them (g5153_chr15_Ss, g5152_chr15_Ss, g5155_chr15_Ss, g3771_chr09_Ss, g4550_chr12_Ss, g3890_chr10_Ss) are also expressed at 5 DAI, 118 encode proteins of unknown function, and 38 encode proteins of the secretome. The GO terms enrichment of this set of genes revealed that extracellular region is the prevalent term (Figure 9). The set of 132 fungal genes particularly expressed in sugarcane are certainly related to host interaction and may contain effectors associated with this singular interaction. Yet, the presence of one gene expressed particularly at 200 DAI encoding a secreted cysteine-protease inhibitor (g2337_chr05_Ss), may be related to the fungal defense against plant proteases. Eight of these genes have homologues in PHI-base, strengthening its involvement in *S. scitamineum* pathogenicity, for instance, g5161_chr15_Ss (PHI:932), g2659_chr06_Ss (PHI:910), g3271_chr08_Ss (PHI:23) which mutants in their orthologs in *U maydis* have reduced pathogenicity, and g672_chr01_Ss (PHI:907) led to loss of pathogenicity (Gold *et al.*, 1994; Kämper *et al.*, 2006).

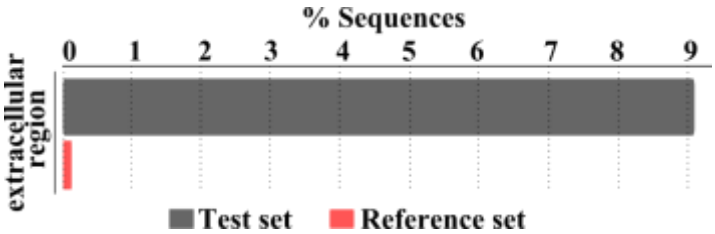


Figure 9. Blast2GO analysis of *S. scitamineum* genes expressed only during interaction. Extracellular region was the only GO term enriched (p-value 0.05).

The genome context of genes expressed particularly *in planta* was analyzed and their distribution revealed the presence of putative effector islands in chromosomes 2, 6, 10, 11, 12, 14, 15 and 16 (Figure 10). Most of the genes are of uncharacterized function and encode secreted proteins having in between 114 to 1257 amino acids. Mig1 related secreted effectors are encoded by genes present in the island chromosome 10 and effectors of the protein family Eff1 are encoded by genes present in the island of chromosome 11. The involvement of the genes in host specificity is strengthened by the evidence of its presence only in the *S. scitamineum* genome, since orthologous for 20 (17%) of them were not found in the genome of its most related species *S. reilianum*, as well as in *U. maydis* and *S. hordei,* according to OrthoMCL analysis.

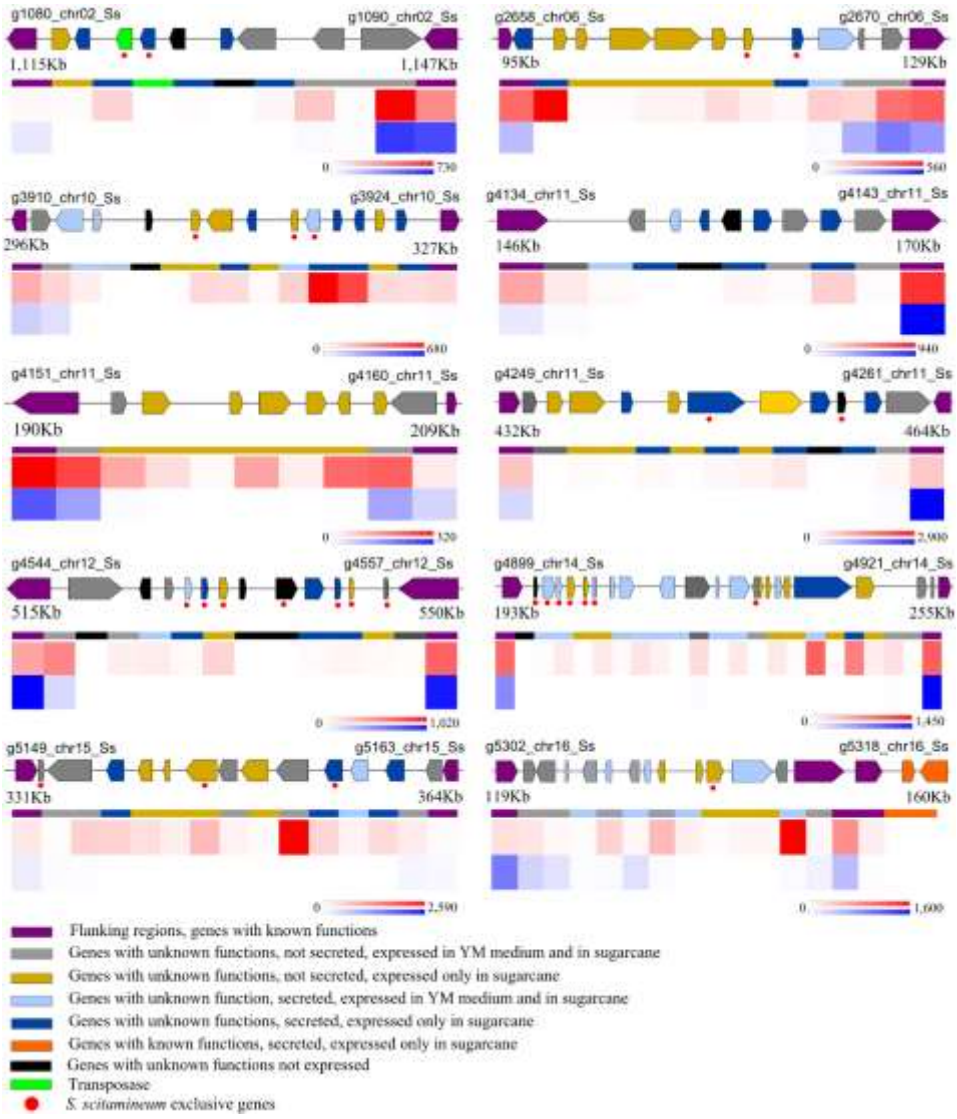


Figure 10. Segments of chromosomes representing the organization of genes in islands (color coded arrows and beneath bars). Expression at 200 DAI (heat map red scale) and *in vitro* (heat map blue scale) are compared using the normalized number of mapped reads, represented by the scales under each chromosome island. Gene names are presented at the borders of each segment of the chromosome, numbers represent the coordinates of these islands in Kb and red dots represents singlets as defined by OrthoMCL (Taniguti *et al*., 2015).

2.4. Discussion

The combination of genome sequencing and transcriptome profiling is a proven strategy to bring insights into pathogen mechanisms to invade host tissues, strategies of acquiring nutrients, avoid plant defense and to provoke disease symptoms. All these events are accomplished by a series of signals inducing a transcriptional reprogramming of its metabolism resulting in survival and dissemination of the pathogen. Even thought at 5 DAI *in planta* a small percentage of fungal genes were detected as transcriptionally active, we detected genes expressed related to initial phases of infection to surpass the physical barriers of plant cell wall, as well as

genes that probably act as pathogenicity-virulence factors or effectors. One of these is codified by g3970_chr10_Ss, which was the most expressed gene of the secretome at 5 DAI. Although encoding an uncharacterized protein, it is a homolog to a protein coding gene of *U. maydis (um03274)*. In the corn smut fungus, its expression was detected only *in planta* and not in axenic cultures (Donaldson *et al.*, 2013b). In *S. scitamineum* the expression of this gene is low during *in vitro* growth. Additional analysis of the sequence revealed the existence of repeats rich in residues proline and glutamine. The function repeats rich in proline and glutamine was vastly described in the human fungal pathogen *Candida albicans*. The 10-amino-acid long N-terminal repeat in the *Hwp1p* adhesin allows covalent cross-linking to host cells (Levdansky *et al.*, 2008; Padovan *et al.*, 2009). Another preferentially expressed gene *in planta,* both at 5 and 200 DAI, codify to a small secreted protein with 135 amino acids (g3890_chr10_Ss), which is specific to the plant interaction, have a homolog in *U. maydis* genome (um03203.1). The genes g3870_chr10_Ss, g488_chr01_Ss and g5684_chr17_Ss were highly expressed in the final phase of the infection cycle and they encode uncharacterized secreted proteins. In these cases, although conserved motifs were not detected, the percentage of glycine residues is high in both proteins. In *M. oryzae*, members of the Pwl gene family codify to small glycine-rich secreted proteins acting as Avrs conferring host specificity (Sweigard *et al.*, 1995). All these genes of undetermined functions but encoding secreted proteins and transcripts detected *in planta* only are good targets for experimental analyses to elucidate potential involvement in fungal growth and disease development.

The analysis of RNAseq data from *S. scitamineum* growth in sugarcane at two different moments in comparison with fungal transcriptome in culture medium showed that approximately 13.5% of its predicted genes are differentially expressed *in planta* considering the early and late moments of the interaction. These genes are related to several metabolic processes important for pathogen spread in the host tissues. One of these processes involves the chitin modification, mechanism that prevents the generation of elicitor active chitin oligomers which would reveal the presence of the pathogen in the plant, triggering defense responses. The deacetylation of surface-exposed chitin into chitosan acts as a molecular disguise strategy, and, consequently, chitin deacetylases are important pathogenicity factors (Nampally *et al.*, 2012). The up-regulation of chitin deacetylase during plant interaction was described in several pathogens such as the wheat pathogen fungus *Puccinia striiformis* f.sp. *tritici* (*Pst*), the cacao hemibiotrophic pathogen *Moniliophthora roreri* (Meinhardt *et al.*, 2014), and in the necrotrophic fungus *Botrytis cinerea* (Leroch *et al.*, 2013). Deacetylase is one of the most up-regulated gene at 5 DAI, indicating that *S. scitamineum* uses this strategy to dodge the plant defense in the early phases of disease

development. Another defense strategy used by the fungus to minimize the plant response is the secretion of a cysteine-protease inhibitor, since the proteolytic machinery of plants plays important roles in defense against pathogens. This protective system was described in the oomycete *Phytophthora infestans*, the agent of the late blight disease of tomato (*Lycopersicon esculentum*) and potato (*Solanum tuberosum*). The *P. infestans* pathogen has evolved an arsenal of proteases inhibitors to overcome the action of hosts proteases, including serine and cysteine proteases inhibitors, that can bind and inhibit pathogenesis-related subtilisin-like serine protease of tomato (Tian *et al.*, 2004).

Another pathogen protective strategy is the ability to detoxify the environment. Plants secrete various antimicrobial compounds into the apoplast to restrict pathogen growth. Examples are steroidal glycoalkaloids, such as saponin, and plant derived reactive oxygen species (ROS), which accumulate upon MAMP perception (Ökmen and Doehlemann, 2014). One of these detoxifying enzymes is pisatin demethylase (g4198_chr11_Ss), which is up-regulated at 200 DAI in *S. scitamineum* transcriptome. The pea pathogen *Fusarium oxysporum f. sp. pisi* is able to detoxify the phytoalexin pisatin, a substrate-inducible cytochrome P450, produced as a defense response by the plant (Coleman *et al.*, 2011). Other genes related to Cytochrome P450 (g1549_chr03_Ss) and benzoate 4-monooxygenase (g4198_chr11_Ss) are up-regulated at 200 DAI in *S. scitamineum* transcriptome, which could also be related to detoxification. This enzyme produce phenolic derivatives that are channeled to the b-ketoadipate pathway for aromatic compound degradation (Harwood and Parales, 1996), being important in plant pathogenic fungi for detoxification of plant metabolites such as benzoic acid and isoeugenol (Podobnik *et al.*, 2008). Catalases, highly relevant to fungal pathogen virulence (Roetzer *et al.*, 2011) is an enzyme involved in oxidative stress response against hydrogen peroxide (H_2O_2) generated by the host plant during the fungal infection, had its transcriptional up-regulated at 5 and 200 DAI in *S. sporisorium* interaction with sugarcane. In this way, superoxide dismutase was one of the most expressed gene at 5 and 200 DAI. Its importance in the initial host penetration is well documented (Weßling *et al.*, 2012), and expression also in the final stages of smut colonization in sugarcane provide an additional protection against oxidative stress. These results show that this *S. scitamineum* isolated has an efficient mechanism scavenging ROS generated by the plant and this feature can contribute to host susceptibility.

During the co-evolution of fungal plant pathogens and their hosts there has been a seesawing interplay between pathogen virulence and host resistance. Thus, to facilitate infection, plant pathogens secrete numerous effector proteins into the plant apoplast or cytosol (Koeck *et al.*, 2011). Besides the strategies used to defend itself from plant immune system, *S. scitamineum*

seems to have an arsenal of effectors that can potentially manipulate host metabolism. In many plant pathogenic microbes, effectors show common features. They are small proteins, potentially secreted, generally cysteine-rich and usually have no homology to known proteins in databases (Stergiopoulos and de Wit, 2009). Additionally, they present similar expression patterns, namely no expression or low-level expression during axenic cultures compared to strong induction of expression during host infection (Soyer *et al.*, 2014). The analysis of differentially expressed genes gave us indications of putative *S. scitamineum* effectors transcriptionally active. One of them a chorismate mutase (g6307_chr21_Ss), up-regulated at 200 DAI, involved in attenuating plant salicylic acid level described in *U. maydis* (Djamei *et al.*, 2011) together with the salicylate hydroxylase (g4103_chr11_Ss). Additionally, the secreted fungal effector Pep1 is essential for penetration of the host epidermis and establishment of biotrophy in the *U. maydis* (Doehlemann *et al.*, 2008), as well as can act as an apoplastic inhibitor of host peroxidases (Hemetsberger *et al.*, 2012). The Pep1 ortholog of *S. scitamineum* (g1816_chr03_Ss) is differentially expressed at 200 DAI, indicating its possible function in other protective mechanisms in the final stages of disease development. Its expression was not analyzed in *S. scitamineum* before, however recent works suggest that Pep1 is a conserved fungal core effector and might play a fundamental role in virulence of biotrophic smut fungi (Hemetsberger *et al.*, 2012; Que, Xu, *et al.*, 2014).

Of the differentially expressed genes at 5 DAI, nine of them have homologues in the pathogen-host interaction database. Noteworthy is three up-regulated that codify to sugar/glucose transporter and maltose permease, that in *U. maydis* mutants shows reduced virulence (Wahl *et al.*, 2010). Among the differentially expressed genes at 200 DAI, 33 of them have homologues in the pathogen-host interaction database. These genes are related to sugar transporter, nicotinic acid transporter, peptide transporter and the secreted proteins beta-glucosidase, lipase and aspartic protease. The sugar transporter codified by g1034_chr02_Ss corresponds to the *U. maydis* plasma membrane-localized sucrose transporter (Srt1), which is sucrose specific, and allows the direct utilization of sucrose without the production of extracellular monosaccharides known to elicit plant immune responses, being considered a fungal virulence factor (Wahl *et al.*, 2010). All these genes probably act as important virulence factors in *S. scitamineum* during all phases of its interaction with sugarcane, since its function in pathogenesis was revealed in other pathogen-host interactions.

As mentioned before the ability to pass through the plant cell wall by secreting of a complex of extracellular cell wall degrading enzymes is evident in *S. scitamineum*. The transcriptome data revealed several genes related to plant cell wall breakdown that are up-regulated at 5 and 200 DAI. Despite the fact that biotrophic fungi have a reduced number of

hydrolases, related to the necessity of minimizing host cell wall damage to avoid triggering plant immunity (Duplessis *et al.*, 2011) they are necessary to entry into plant tissue and are up-regulated in several plant-pathogen interactions (Garnica *et al.*, 2013; Kawahara *et al.*, 2012; Meinhardt *et al.*, 2014). The production of laccase as one of the differentially expressed genes in sugarcane also reveals the *S. scitamineum* ability to breakdown lignified tissues. Laccase is a polyphenol oxidase that catalyzes the reduction of O_2 to H_2O using a range of phenolic compounds as hydrogen donors, including the lignin (Thurston, 1994). Lignin is the second most abundant constituent of the vascular plants cell wall, acting in cellulose protection towards hydrolytic microbial attack (Ruiz-Dueñas and Martínez, 2009).

The annotation of *S. scitamineum* genome revealed the presence of three genes that codify to laccases (g1208_chr02_Ss, g3267_chr08_Ss and g4962_chr14_Ss). Due to the properties of its substrate, the enzyme that participate in the breakdown of lignin should be extracellular (Baldrian, 2006). The *S. scitamineum* laccase codified by the gene g1208_chr02_Ss is part of the pathogen secretome and is up-regulated both at 5 and 200 DAI, with values of Log_2FC of 6.56 and 7.59, respectively. In this sense, this enzyme must be involved in lignin breakdown in sugarcane performed by the fungus, being important to pathogen spread as well as has potential to be studied in innumerous biotechnological applications (Madhavi and Lele, 2009; Mayer and Staples, 2002; Singh Arora and Kumar Sharma, 2010).

Laccase is also involved in various relevant physiological processes, including the development of fungal fruit bodies (Leatham and Stahmann, 1981) and pigmentation of fungal spores (Clutterbuck, 1972). In the transcriptome analysis of *S. scitamineum* we found that besides the up-regulation of an extracellular laccase (g1208_chr02_Ss) at 200 DAI, other not secreted laccase is up-regulated (g4962_chr14_Ss, $Log_2FC = 5.49$). In this case, the enzyme is possibly related to pigment formation, once this moment of the interaction is characterized by intensive teliospore differentiation. As well as, at 200 DAI, several polyketide synthases related to pigment biosynthesis are also up-regulated.

Yet to improve further the analysis of *S. scitamineum* transcriptional profiles, we searched the distribution of genes specifically expressed *in planta*, which allowed the identification of 10 putative effector islands in *S. scitamineum* genome. The presence of effector islands is widespread in fungal pathogens genomes. In *U. maydis* genome were found 12 islands of genes encoding small secreted proteins with unknown function, most of them are regulated together and induced in infected tissue, and deletion of individual islands altered the pathogen virulence, leading to a complete lack of symptoms or hypervirulence (Kämper *et al.*, 2006). The *S. scitamineum* effectors island present in chromosome 10 are composed by the Mig1 genes. The family of Mig-1-related

secreted effectors in *U. maydis* are on chromosome 8 in an island of 3 genes, as well as in *S. reilianum*, where the island of 8 Mig1 related proteins are in chromosome 8 (Wollenberg and Schirawski, 2014). The *mig1* gene of *U. maydis* was the first gene identified in this organism whose expression is coupled to the biotrophic phase. Its expression in the maize pathogen is undetectable during hyphal growth on the leaf surface and formation of infection structures but is immediately switched on after penetration and remains high during fungal colonization, however becomes virtually undetectable in mature teliospores (Basse *et al.*, 2002). In contrast, the *S. scitamineum mig1* related genes are expressed until the final phase of fungal cycle *in planta*, when the teliospores formation is at its peak.

The involvement of the genes present in islands in the host specificity is strengthened by the evidence of its presence only in the *S. scitamineum* genome, since orthologous for 20 (17%) of them were not found in the genome of its most related species *S. reilianum*, as well as in *U. maydis* and *S. hordei*. Host specificity is an important trait underlying the interaction of smuts with their hosts, but is still poorly understood at the molecular level. Despite being phylogenetically close, the smut fungi infect different Poaceae, and vary in their mode of plant colonization and symptom development. Searching for species-specific genes is a promising strategy to identify genes involved in host-specific adaptations (Wollenberg and Schirawski, 2014), mainly sets of highly specialized effector proteins that enable the fungal proliferation and, concomitantly, the escape of the plant defense system (Feldbrügge *et al.*, 2013). Another important characteristic found in four predicted islands (chromosomes 2, 6, 10 and 11c) was the presence of repetitive elements, that has been viewed as drivers of genome evolution (Schmidt and Panstruga, 2011), and can be related to its adaptability to sugarcane.

REFERENCES

Agrios, G.N. (2005) Plant pathology, Elsevier Academic Press.

Ailloud, F., Lowe, T.M., Robène, I., Cruveiller, S., Allen, C. and Prior, P. (2016) In planta comparative transcriptomics of host-adapted strains of *Ralstonia solanacearum* Josenhans, C., ed. *PeerJ* **4**, e1549.

Arora, D.S. and Sharma, R.K. (2010) Ligninolytic Fungal Laccases and Their Biotechnological Applications. *Appl. Biochem. Biotechnol.* **160**, 1760–1788.

Bakkeren, G., Kämper, J. and Schirawski, J. (2008) Sex in smut fungi: structure, function and evolution of mating-type complexes. *Fungal Genet. Biol.* **45**, 15–21.

Baldrian, P. (2006) Fungal laccases - occurrence and properties. *FEMS Microbiol. Rev.* **30**, 215–42.

Basse, C.W., Kolb, S. and Kahmann, R. (2002) A maize-specifically expressed gene cluster in *Ustilago maydis*. *Mol. Microbiol.* **43**, 75–93.

Begerow, D., Stoll, M. and Bauer, R. (2006) A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data. *Mycologia* **98**, 906–16.

Carvalho, G., Quecine, M.C., Longatto, D.P., et al. (2016) *Sporisorium scitamineum* colonisation of sugarcane genotypes susceptible and resistant to smut revealed by GFP-tagged strains. *Ann. Appl. Biol.* **169**, 329–341.

Castell-Miller, C. V, Gutierrez-Gonzalez, J.J., Tu, Z.J., Bushley, K.E., Hainaut, M., Henrissat, B. and Samac, D.A. (2016) Genome assembly of the fungus *Cochliobolus miyabeanus*, and transcriptome analysis during early stages of infection on american wildrice (*Zizania palustris* L.) Yu, J.-H., ed. *PLoS One* **11**, e0154122.

Clutterbuck, A.J. (1972) Absence of laccase from yellow-spored mutants of *Aspergillus nidulans*. *J. Gen. Microbiol.* **70**, 423–35.

Coleman, J.J., Wasmann, C.C., Usami, T., White, G.J., Temporini, E.D., McCluskey, K. and VanEtten, H.D. (2011) Characterization of the gene encoding pisatin demethylase (FoPDA1) in *Fusarium oxysporum*. *Mol. Plant. Microbe. Interact.* **24**, 1482–91.

Djamei, A., Schipper, K., Rabe, F., et al. (2011) Metabolic priming by a secreted fungal effector. *Nature* **478**, 395–398.

Doehlemann, G., Wahl, R., Horst, R.J., et al. (2008) Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. *Plant J.* **56**, 181–195.

Donaldson, M.E., Meng, S., Gagarinova, A., Babu, M., Lambie, S.C., Swiadek, A.A. and Saville, B.J. (2013) Investigating the *Ustilago maydis/Zea mays* pathosystem: Transcriptional responses and novel functional aspects of a fungal calcineurin regulatory B subunit. *Fungal Genet. Biol.* **58**, 91–104.

Duplessis, S., Cuomo, C.A., Lin, Y.-C., et al. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9166–71.

Feldbrügge, M., Kellner, R. and Schipper, K. (2013) The biotechnological use and potential of plant pathogenic smut fungi. *Appl. Microbiol. Biotechnol.* **97**, 3253–65.

Garnica, D.P., Upadhyaya, N.M., Dodds, P.N., et al. (2013) Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing Matz, M. V., ed. *PLoS One* **8**, e67150.

Gold, S., Duncan, G., Barrett, K. and Kronstad, J. (1994) cAMP regulates morphogenesis in the fungal pathogen *Ustilago maydis*. *Genes Dev.* **1**, 2805–2816.

Harwood, C.S. and Parales, R.E. (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* **50**, 553–90.

Hemetsberger, C., Herrberger, C., Zechmann, B., et al. (2012) The *Ustilago maydis* effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity Xu, J.-R., ed. *PLoS Pathog.* **8**, e1002684.

Islamovic, E., García-Pedrajas, M.D., Chacko, N., Andrews, D.L., Covert, S.F. and Gold, S.E. (2015) Transcriptome analysis of a *Ustilago maydis ust1* deletion mutant uncovers involvement of laccase and polyketide synthase genes in spore development. *Mol. Plant-Microbe Interact.* **28**, 42–54.

Kämper, J., Kahmann, R., Bölker, M., et al. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101.

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T. and Minami, E. (2012) Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS One* **7**, e49423.

Koeck, M., Hardham, A.R. and Dodds, P.N. (2011) The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cell. Microbiol.* **13**, 1849–57.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.

Leatham, G.F. and Stahmann, M.A. (1981) Studies on the laccase of Lentinus edodes: specificity, localization and association with the development of fruiting bodies. *Microbiology* **125**, 147–157.

Leroch, M., Kleber, A., Silva, E., Coenen, T., Koppenhöfer, D., Shmaryahu, A., Valenzuela, P.D.T. and Hahn, M. (2013) Transcriptome profiling of Botrytis cinerea conidial germination reveals upregulation of infection-related genes during the prepenetration stage. *Eukaryot. Cell* **12**, 614–26.

Levdansky, E., Sharon, H. and Osherov, N. (2008) Coding fungal tandem repeats as generators of fungal diversity. *Fungal Biol. Rev.* **22**, 85–96.

Madhavi, V. and Lele, S.S. (2009) Laccase: properties and applications. *BioResources* **4**, 1694–1717.

Martínez-Espinoza, A.D., García-Pedrajas, M.D. and Gold, S.E. (2002) The Ustilaginales as plant pests and model systems. *Fungal Genet. Biol.* **35**, 1–20.

Mayer, A.M. and Staples, R.C. (2002) Laccase: new functions for an old enzyme. *Phytochemistry* **60**, 551–565.

McGettigan, P.A. (2013) Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* **17**, 4–11.

Meinhardt, L.W., Costa, G.G., Thomazella, D.P., et al. (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.

Morrison, E.N., Donaldson, M.E. and Saville, B.J. (2012) Identification and analysis of genes expressed in the *Ustilago maydis* dikaryon: uncovering a novel class of pathogenesis genes. *Can. J. plant Pathol.* **34**, 417–435.

Nampally, M., Moerschbacher, B.M. and Kolkenbrock, S. (2012) Fusion of a novel genetically engineered chitosan affinity protein and green fluorescent protein for specific detection of chitosan in vitro and in situ. *Appl. Environ. Microbiol.* **78**, 3114–9.

Ökmen, B. and Doehlemann, G. (2014) Inside plant: biotrophic strategies to modulate host immunity and metabolism. *Curr. Opin. Plant Biol.* **20**, 19–25.

Padovan, A.C.B., Chaves, G.M., Colombo, A.L. and Briones, M.R.S. (2009) A novel allele of HWP1, isolated from a clinical strain of *Candida albicans* with defective hyphal growth and biofilm formation, has deletions of Gln/Pro and Ser/Thr repeats involved in cellular adhesion. *Med. Mycol.* **47**, 824–35.

Palhares, A.C. (2014) *Análise, via RNAseq, do transcritoma da cana-de-açúcar e identificação de genes expressos em resposta a Sporisorium scitamineum, o agente causal do carvão*. Piracicaba: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo.

Podobnik, B., Stojan, J., Lah, L., Krasevec, N., Seliskar, M., Rizner, T.L., Rozman, D. and Komel, R. (2008) CYP53A15 of *Cochliobolus lunatus*, a target for natural antifungal compounds. *J. Med. Chem.* **51**, 3480–6.

Que, Y., Xu, L., Wu, Q., et al. (2014) Genome sequencing of *Sporisorium scitamineum* provides insights into the pathogenic mechanisms of sugarcane smut. *BMC Genomics* **15**, 1–19.

Rabe, F., Ajami-Rashidi, Z., Doehlemann, G., Kahmann, R. and Djamei, A. (2013) Degradation of the plant defence hormone salicylic acid by the biotrophic fungus *Ustilago maydis*. *Mol. Microbiol.* **89**, 179–188.

Roetzer, A., Gabaldón, T. and Schüller, C. (2011) From *Saccharomyces cerevisiae* to *Candida glabratain* a few easy steps: important adaptations for an opportunistic pathogen. *FEMS Microbiol. Lett.* **314**, 1–9.

Ruiz-Dueñas, F.J. and Martínez, A.T. (2009) Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* **2**, 164–77.

Schmidt, S.M. and Panstruga, R. (2011) Pathogenomics of fungal plant parasites: what have we learnt about pathogenesis? *Curr. Opin. Plant Biol.* **14**, 392–9.

Soyer, J.L., Ghalid, M. El, Glaser, N., et al. (2014) Epigenetic control of effector gene expression in the plant pathogenic fungus *Leptosphaeria maculans* Talbot, N.J., ed. *PLoS Genet.* **10**, e1004227.

Stergiopoulos, I. and Wit, P.J.G.M. de (2009) Fungal Effector Proteins. *Annu. Rev. Phytopathol.* **47**, 233–263.

Stoll, M., Begerow, D. and Oberwinkler, F. (2005) Molecular phylogeny of *Ustilago*, *Sporisorium*, and related taxa based on combined analyses of rDNA sequences. *Mycol. Res.* **109**, 342–56.

Sundar, A.R., Ashwin, N.M.R., Barnabas, E.L., Malathi, P. and Viswanathan, R. (2015) Disease resistance in sugarcane - an overview. *Sci. Agrar. Parana.* **14**, 200–212.

Sweigard, J.A., Carroll, A.M., Kang, S., Farrall, L., Chumley, F.G. and Valent, B. (1995) Identification, cloning, and characterization of PWL2, a gene for host species specificity in the rice blast fungus. *Plant Cell* **7**, 1221–33.

Taniguti, L.M., Schaker, P.D.C., Benevenuto, J., et al. (2015) Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with Sugarcane. *PLoS One* **10**, e0129318.

Thurston, C.F. (1994) The structure and function of fungal laccases. *Microbiology* **140**, 19–26.

Tian, M., Huitema, E., Cunha, L. Da, Torto-Alalibo, T. and Kamoun, S. (2004) A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. *J. Biol. Chem.* **279**, 26370–7.

Wahl, R., Wippel, K., Goos, S., Kämper, J. and Sauer, N. (2010) A novel high-afinity sucrose transporter is required for virulence of the plant pathogen *Ustilago maydis* Heitman, J., ed. *PLoS Biol.* **8**, e1000303.

Weßling, R., Schmidt, S.M., Micali, C.O., Knaust, F., Reinhardt, R., Neumann, U., Loren van Themaat, E. Ver and Panstruga, R. (2012) Transcriptome analysis of enriched *Golovinomyces orontii* haustoria by deep 454 pyrosequencing. *Fungal Genet. Biol.* **49**, 470–82.

Wollenberg, T. and Schirawski, J. (2014) Comparative genomics of plant fungal pathogens: the *Ustilago-Sporisorium* paradigm. *PLoS Pathog.* **10**, e1004218.

Zahiri, A.R., Babu, M.R. and Saville, B.J. (2005) Differential gene expression during teliospore germination in *Ustilago maydis*. *Mol. Genet. Genomics* **273**, 394–403.

3. PLANT CELL LOCALIZATION OF CANDIDATE EFFECTOR PROTEINS FROM *SPORISORIUM SCITAMINEUM*, THE SUGARCANE SMUT PATHOGEN

Abstract

The analysis of *Sporisorium scitamineum* transcriptional reprogramming related to sugarcane colonization unreveal a repertoire of genes encoding uncharacterized proteins, potentially secreted by the pathogen and among the most expressed in early interaction. These genes are good targets to functional analysis aiming to describe their function and its relationship with plant susceptibility. The experiments described in this chapter were developed at the University of California, Davis, under the advice of Prof. Dr. Savithramma Dinesh-Kumar for 5 months (BEPE – FAPESP scholarship, Process 2016/04429-0). The objective was to implement a system to study pathogen effectors in the model plant *Nicotiana benthamiana*. Transient expression technique to describe plant cell compartments targeted by *S. scitamineum* secreted proteins, was used. Results revealed that genes most expressed by the pathogen in early interaction encode proteins which target various plant compartments, including nucleus, and seems to undergo post-translational modifications. Future perspectives include identification of plant proteins interacting with such effectors, which will allow to build up hypothesis about host pathways hijacked by the pathogen; and obtain transgenic plants expressing putative effectors to evaluate phenotypic alterations related to effector expression independent of the pathogen presence.

3.1. Introduction

Effectors are molecules secreted by plant-associated organisms that act in host's apoplast or cytoplasm compartments and may contribute to plant resistance or susceptibility, depending on the host genetic background (Hogenhout *et al.*, 2009; Win *et al.*, 2012). They have evolved to enable parasitism, modulating host physiology – the so called *effector-triggered susceptibility* (ETS), often through suppression of plant immunity or protecting the pathogen from host defense (De Jonge *et al.*, 2011; Win *et al.*, 2012).

Effectors are extraordinary examples of biological innovation. Their functions in host are extremely diverse, targeting different subcellular compartments (Petre and Kamoun, 2014), associate with multiple plant targets or affect distinct processes in the host plant (Kmen and

Doehlemann, 2014; Win *et al.*, 2012) and also showing organ-specific expression (Skibbe *et al.*, 2010). In the apoplast, effectors may act as cell wall degrading enzymes, inhibit extracellular host proteases (van Esse *et al.*, 2008; De Jonge *et al.*, 2011; Shabab *et al.*, 2008; Song *et al.*, 2009), interfere in chitin perception (van den Burg *et al.*, 2006; de Jonge *et al.*, 2010; de Jonge and Thomma, 2009; Takahara *et al.*, 2016), contribute to detoxification (Bouarab *et al.*, 2002; Ökmen *et al.*, 2013) and block peroxidase driven oxidative burst (Hemetsberger *et al.*, 2012). Effectors can also be translocated to host cytoplasm, where they can perform a broad range of activities. In fungal pathogens, the translocation process is still under investigation, but seems to be associated to exocytosis of Golgi-derived secretory vesicles (Panstruga and Dodds, 2009), and the development of a biotrophic interfacial complex (BIC) where effectors accumulate (Giraldo *et al.*, 2013; Khang *et al.*, 2010) or may involve a lipid raft-mediated endocytosis (Gan *et al.*, 2010; Kale *et al.*, 2010). Also, sets of effectors may be delivered in a coordinated manner as pathogenesis progresses (Panstruga and Dodds, 2009). Inside host cells fungal effectors function has been elucidated, for example in suppressing the resistance mediated by R proteins (Houterman *et al.*, 2008), interfering in SA defense-related signalization (Caillaud *et al.*, 2013a; Djamei *et al.*, 2011), regulating anthocyanin biosynthesis to compete with lignification (Tanaka *et al.*, 2014) and manipulating host nutrient efflux to redirect sugar flux to support pathogen propagation, for example, in the *U. maydis* expression of a plasma membrane localized sucrose transporter (Srt1) enable an efficient carbon supply for the fungus and reduces apoplastic sugar elicitors that could trigger plant defenses (Wahl *et al.*, 2010).

However, effectors can also activate plant immune response when recognized by plant resistance gene analogs (RGAs). These genes have conserved domains and motifs that play specific roles in pathogen resistance. For instance, RGAs can be grouped in either transmembrane leucine rich repeat (TM-LRR) or nucleotide binding site leucine rich repeat (NBS-LRR) (Sekhwal *et al.*, 2015). The TM-LRR can be receptor like kinases (RLKs) acting as pattern recognition receptors (PRRs), which recognize pathogen/microbe associated molecular pattern (PAMP/MAMP) and trigger immunity (PTI/MTI) to a wide range of pathogens. The NBS-LRRs direct (gene-to-gene model) or indirect (guard, decoy and bait models) recognize specific virulence proteins inside the host cell, which leads to the effector triggered immunity (ETI) plant response (Caplan *et al.*, 2008; Lo Presti *et al.*, 2015; Win *et al.*, 2012), resulting in extensive transcriptional reprogramming of the host.

Pathogen effector catalogs are highly lineage-specific and determination of effector catalogs is a challenge. Typical effector calling based on the presence of signal peptides and absence of transmembrane domains, but major challenge for the future will be to assign

biological functions to the increasing number of effector molecules identified in fungal genomes (De Jong*e et al*, 2011). RNAseq data from smut infected plants associated with genomic data were used in *S. scitamineum* – sugarcane interaction to predict pathogen genes preferentially expressed in early and late interactions, as well as differentially expressed genes in comparison to its axenic growth in culture medium and genes expressed only during interaction (Chapter 1, Taniguti *et al.*, 2015). Early after inoculation (5 DAI), the third most expressed gene encode to a secreted protein (g3970_chr10_Ss) that presents several repeats, such as the "PQPQDGQ" motif represented seven times close to the N-terminal region and "PYGDKPNGDAENSDS" repeated eight times towards the C-terminal region. Other four preferentially expressed genes (g2_chr01_Ss, g3890_chr10, g6307_chr21_Ss and g1513_chr03_Ss) encode small secreted proteins of 236, 135, 321 and 215 amino acids respectively with no identifiable conserved domains or any sequence feature. Additionally, g3890_chr10_Ss expression was detected only *in planta*. Genes expressed only after whip development were 131, and six of them (g5153_chr15_Ss, g5152_chr15_Ss, g5155_chr15_Ss, g3771_chr09_Ss, g4550_chr12_Ss, g3890_chr10_Ss) were also expressed at 5 DAI, 118 encoded proteins of unknown function, and 38 encoded proteins of the secretome (Taniguti *et al.*, 2015).

These genes are candidates to functional analysis aiming to understand mechanisms associated to pathogenicity and to improve our knowledge to protect sugarcane crops from disease development (Petre *et al.*, 2015). Effectors emerged as tools in disease resistance breeding by accelerating the identification and functional characterization of host resistance genes (Vleeshouwers and Oliver, 2014), and definition of a effector exact location into the host cellular compartment can be the first step in the approach to identify the receptor protein. With the aim to identify plant cell compartments targeted by *S. scitamineum* effector proteins, we performed experiments of transient expression in *N. benthamiana* associated to confocal microscopy and immunoblots.

3.2. Material and methods

3.2.1. Effectors selection

Were selected four *S. scitamineum* genes among the preferentially expressed at 5 DAI (Taniguti et al., 2015). g3970_chr10 is the most expressed gene of *S. scitamineum* secretome at 5 DAI according to RNAseq analysis (Chapter 1), codifying to a 745 aa protein. The other three selected genes were g2_chr01, g3890_chr10 and g1513_chr03, which encode to small secreted

proteins of 236, 135 and 215 amino acids respectively. These genes have no identifiable conserved domains or any sequence feature.

3.2.2. Primer design and amplifications

Amplifications of target genes were made in a two-step approach to get *attB* Gateway recombination sites. Primers were manually designed and tested for secondary structures formation using NetPrimer (http://www.premierbiosoft.com/netprimer/). First PCR was made using primers containing 18-20 nucleotides of the open reading frame (ORF) coding the mature form of the effector protein (i.e., without the signal peptide and stop codon sequences) and 10 nucleotides of *attB1* and *attB2* sequences (for forward and reverse primers, respectively), as described in Table 5. cDNA samples prepared from sugarcane buds 5 day after inoculation with *S. scitamineum* 39 teliospores using punction method were used in amplifications. RNA was extracted using Trizol® and cDNA prepared using Superscript II RT kit (Invitrogen), according to manufacturer's instructions. PCRs were made using KAPA HiFi HotStart PCR Kit (Kapa Biosystems) and were composed by 1X KAPA HiFi Fidelity Buffer, 0.3 mM dNTP mix, 0.3 µM of each primer (forward and reverse), 25 ng of cDNA and 1 U Kapa HiFi HotStart DNA polymerase. Reactions were carried out using the following cycling (Veriti, Applied Biosystems): initial denaturation 95ºC for 3 min, 35 cycles of denaturation at 98ºC for 20 sec, annealing at 67ºC for 30 sec, and extension at 72ºC for 60 sec/Kb, and final extension for 2 min at 72ºC. The amplification product was then used in a second PCR with *attB1* (5' GGGGACAAGTTTGTACAAAAAAGCAGGCT 3') and *attB2* (5' GGGGACCACTTTGTACAAGAAAGCTGGGT 3') primers. Reaction was composed by 1X KAPA HiFi Fidelity Buffer, 0.3 mM dNTP mix, 0.3 µM of each primer (forward and reverse), 10 µL of the first reaction and 1 U Kapa HiFi HotStart DNA polymerase. Cycling was composed by an initial denaturation at 95ºC for 1 min, 5 cycles of denaturation at 98ºC for 20 sec, annealing at 55ºC for 30 sec, and extension at 72ºC for 30 sec/Kb, 20 cycles of denaturation at 98ºC for 20 sec, annealing at 64ºC for 30 sec, and extension at 72ºC for 30 sec/Kb and final extension at 72ºC for 1 min. First and second PCRs amplifications were confirmed in 1% agarose gels using 1 Kb (Thermo Scientific) as ladder and SyBr green as staining.

Table 5. Primers designed to get amplicons containing *attB1* sites for obtain Gateway ENTRY vectors.

Gene	Primer Forward	Primer Reverse
g2_chr01	AGCAGGCTTCACCATGCTGCAGTGCAGCC TGTCGAC	GAAAGCTGGGTCCTTCAAAGCGGCCTG GTAAAGGA
g3890_chr10	AGCAGGCTTCACCATGACGATCGGCCGTG CGGGT	GAAAGCTGGGTCGTTGCCACCCTTGGG CTTC
g1513_chr03	AGCAGGCTTCACCATGCGAGTCATCGACA AGCTCT	GAAAGCTGGGTCAGGCAGTATCTCAGG CTTGA
g3970_chr10	AGCAGGCTTCACCATGACCCCCGCCATGG CCAACA	GAAAGCTGGGTCCTGCAAGTAGTCCTCC TGCT
RED = partial *attB* sequence Blue = Kozac sequence Green = Start codon Black = effector sequence		

3.2.3. Plasmids and cloning procedures

PCR products containing *attB* sites were purified (GFX™ PCR DNA and Gel Band Purification Kit, Sigma Aldrich) and recombined in *pENTR221* plasmid using Gateway® BP Clonase® II Enzyme mix (Sigma-Aldrich) according to manufacturer's instructions. *Escherichia coli* DH5α quimiocompetent cells were transformed with the recombinant plasmids, and grown in selective medium containing kanamycin (50 µg/mL). Grown colonies were selected and multiplied plasmids were extracted using QIAprep Spin Miniprep Kit (Quiagen). Inserts were sequenced using M13F (5' GTAAAACGACGGCCAGT 3') and M13R (5' AACAGCTATGACCATG 3') primers to confirm sequence. Following, LR recombination in a destination vector containing 35S promoter and 3xHA/4xMyc/Citrine tags were made using Gateway® LR Clonase® II Enzyme mix (Sigma-Aldrich) according to manufacturer's instructions. Recombinant plasmids were used to transform quimiocompetent *E. coli* DH10B cells. Colonies grown in solid LB medium containing streptomycin (50 µg/mL) were multiplied, the plasmids extracted and sequence confirmed using Sanger sequencing. Confirmed vectors were transformed into *A. tumefaciens* strain GV2260. All transformed bacteria were be conserved at -80ºC in 20% glycerol.

3.2.4. Agroinfiltration

N. benthamiana plants were cultivated in greenhouse conditions at 22ºC under photoperiod on time intervals of 16-h day and 8-h night. *A. tumefaciens* GV2260 was used to deliver T-DNA constructs into leaf cells of 3-week-old *N. benthamiana* plants, following the agroinfiltration method previously described (WIN et al. 2011). Briefly, overnight-grown bacterial cultures were adjusted to an optical density at 600 nm (OD600) of 0.1 into an infiltration buffer

(10 mM $MgCl_2$, 10 mM MES, 200 μM acetosyringone). The leaves were collected 2 days after infiltration for further microscopy and protein extraction.

3.2.5. Live-cell imaging by laser-scanning confocal microscopy

Small pieces of leaves were mounted in water between a slide and a coverslip (inferior face toward the objective) and immediately observed. Live-cell imaging was performed with a Zeiss LSM710 confocal microscope equipped with a LDC-apochromat 40×/1.1W Korr M27. Excitation laser wavelength was 514 nm and emission 543 nm.

3.2.6. Total protein isolation

N. benthamiana leaves were harvested 2 days after infiltration, frozen in liquid nitrogen, and ground into powder with mortar and pestle. Total protein extraction was performed by adding extraction buffer (50 mM HEPES pH 7.3, 15 mM EDTA pH 8.0, 5% (v) glycerol, 250 mM sucrose, 3 mM DTT, 1x Protease Inhibitor Cocktail (Sigma), 0.002 vol of IGEPAL® CA-630). Samples were centrifuged (10 min, 14,000 rpm, 4ºC) and the supernatant transferred to a new tube. To each sample were added 70 μl of 5X Protein Loading Buffer (National Diagnosis). Proteins were denatured by heating in water bath for 5 min. Then, 10 to 40 μl of isolated proteins were separated by 12% SDS-PAGE.

3.2.7. Nuclear protein extraction

For enrichment of nuclear proteins, to 500 mg of grinded tissue were added 5 ml of extraction buffer 1 (0.4 M sucrose, 10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.1 mM PMSF, cOmplete™ Protease Inhibitor Cocktail - Sigma), and samples were kept in ice under agitation for 20 min. Then the solution was filtered (Falcon® 100 μm Cell Strainer) and centrifuged (25 min, 5,000 rpm, 4ºC). The pellet was resuspended in 1 ml of extraction buffer 2 (0.25 M sucrose, 10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1% (v/v) Triton X-100, 5 mM β-mercaptoethanol, 0.1 mM PMSF, and protease inhibitors tablets) and centrifuged (10 min, 12,000 rpm, 4ºC). The pellet was resuspended in 350 μl of extraction buffer 3 (1.7 M sucrose, 10 mM Tris-HCl pH 8.0, 2 mM $MgCl_2$, 0.15% (v/v) Triton X-100, 5 mM β-mercaptoethanol, 0.1 mM PMSF, and protease inhibitors tablets) and the solution layered onto 350 μl of extraction buffer 3. Samples were centrifuged (60 min, 13,000 rpm, 4ºC). The pellet was

resuspended in 200 µl of nuclei lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 2% SDS, and protease inhibitors tablets) and added 50 µl of 5X Protein Loading Buffer (National Diagnosis). Proteins were denatured by heating in water bath for 5 min. Then, 10 to 40 µl of isolated proteins were separated by 12% SDS-PAGE.

3.2.8. Immunoblot analyses

Immunoblots aimed to check the integrity of the fusion proteins present in total/nuclear extracts. Proteins separated by 12% SDS-PAGE were electrotransferred onto a polyvinylidene diflouride membrane using a Trans-Blot turbo transfer system (Bio-Rad, Munich). The membrane was blocked with 5% bovine serum albumin (BSA) in 1x PBS and Tween 20 0.1%. GFP detection was performed using with a rat anti-GFP 5F8 antibody (Chromotek, Munich) and a HRP-conjugated antirat antibody. Membrane revelation was carried out with an ImageQuant LAS 4000 luminescent imager (GE Healthcare Life Sciences, Piscataway, NJ, U.S.A.).

3.3. Results

The selection of *S. scitamineum* genes was based on the previous results obtained in RNAseq experiments *in vitro* and *in planta* pathogen growth (Chapter 1, Taniguti et al., 2015). To determine where the 4 candidate effectors accumulate in plant cells, we cloned the coding sequence matching their mature form (i.e., without signal peptide) to obtain candidate effector-green fluorescent protein (Citrine) fusions downstream of a 35S promoter in an *Agrobacterium tumefaciens* binary vector. Then, we expressed the fusion proteins in *N. benthamiana* by agroinfiltration and determined their accumulation in leaf cells by confocal microscopy. All proteins accumulated at detectable levels in leaves (Figure 11) and the possible localization was determined. Protein G2_chr01 was accumulated in cytosol and nucleus, while G1513_chr03 localizes in membrane and nucleus. G3890_chr10 expression was detected in nucleus and in some vesicles, but nuclear localization was predominant. G3970_chr10 expression in nucleus was not detected, but it seems to surround nucleus, for this reason one hypothesis is that it may be targeted to endoplasmic reticulum (ER) and cytosol.

Western blot analysis was performed to check integrity of Citrine fused effectors. First, we used a total protein extraction method, which allowed the detection of G2_chr01, G1513_chr03 and G3970_chr10 fusioned proteins with the expected sizes (Figure 12A). However, the predominance of a second band for G3970_chr10 higher than the expected size

(Figure 12A) indicates the occurrence of post-translational events. G3890_chr10 was not detected in protein extracts using this protein extraction protocol. Considering that in transient expression analysis G3890_chr10 detection was predominant in nucleus, we used a protein extraction method to enrichment of nuclear proteins, allowing its detection in the right size (Figure 12B), reinforcing the hypothesis that this effector target plant nucleus.

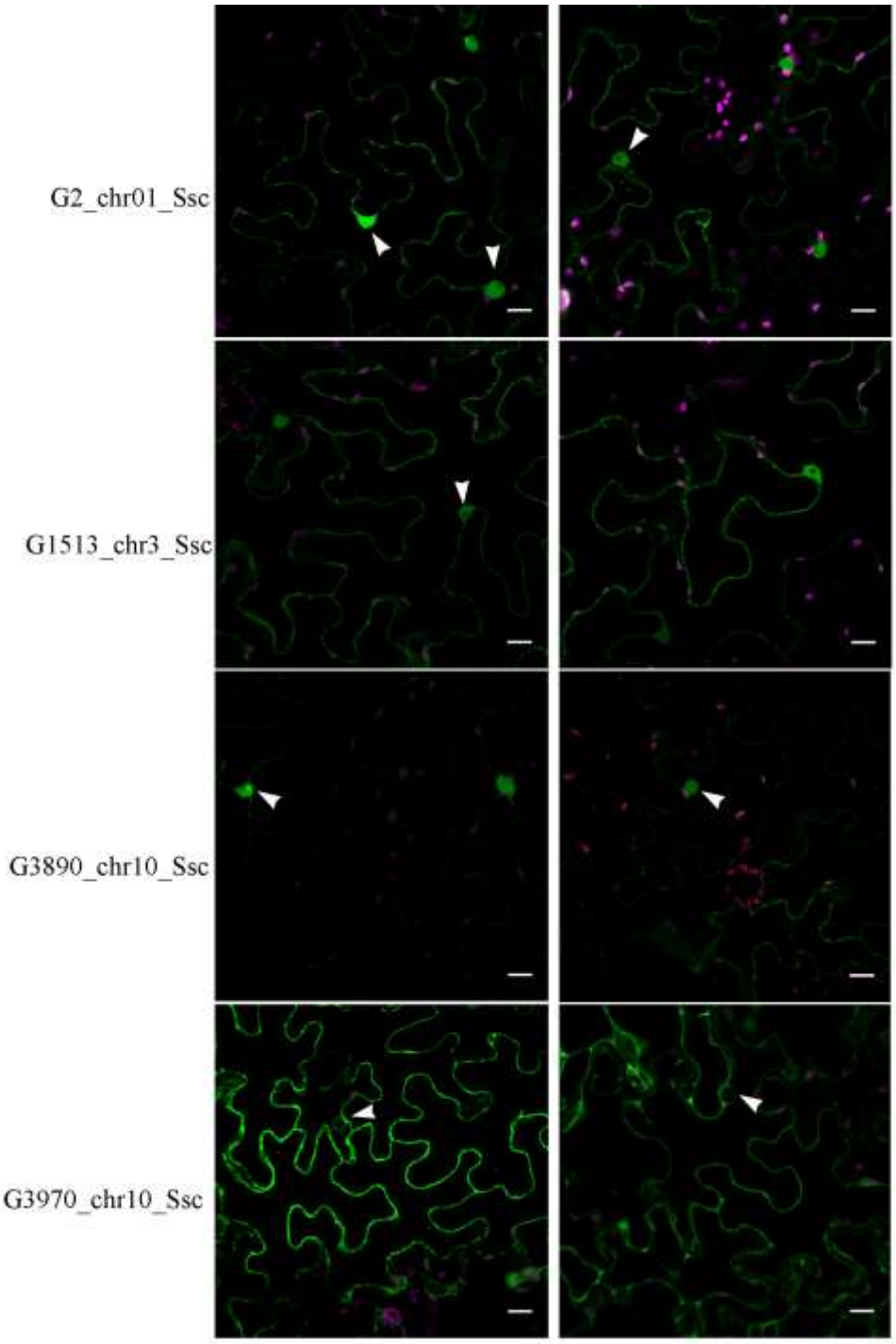


Figure 11. *S. scitamineum* candidate effectors accumulate in distinct subcellular compartments. Fusion proteins were transiently expressed in *Nicotiana benthamiana* leaf cells by agroinfiltration. Live-cell imaging was performed with a laser-scanning confocal microscope 2 days after infiltration.

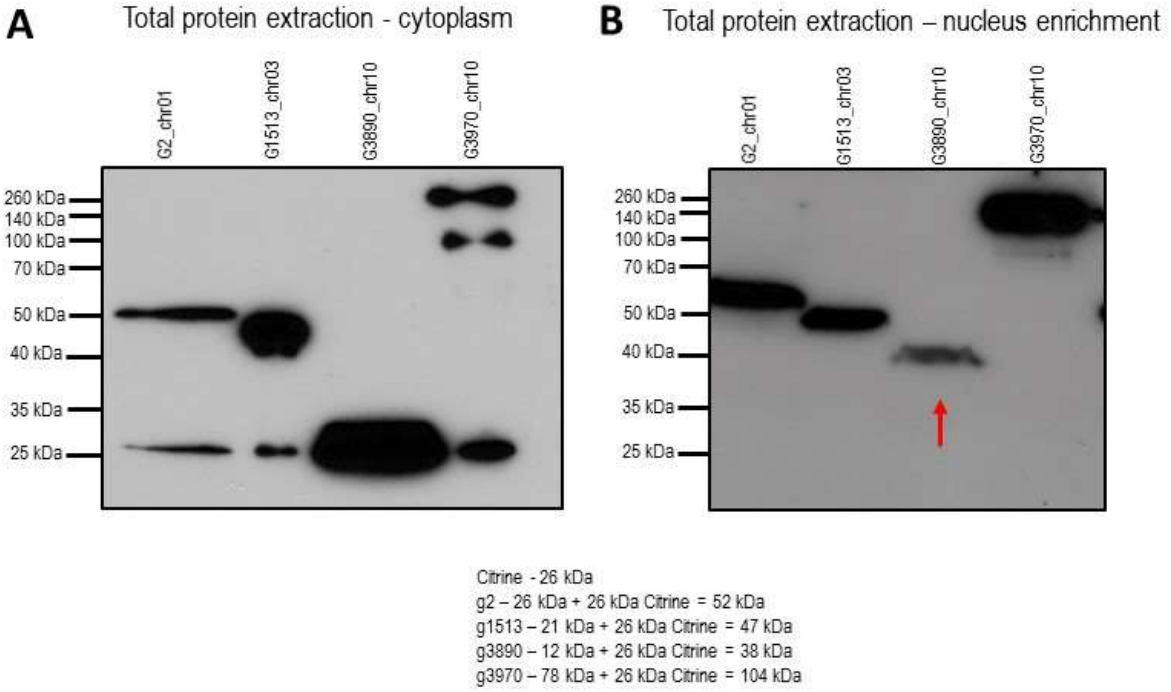


Figure 12. Western blots performed with total protein extracts (A) or nuclear proteins enriched extracts (B). Proteins were isolated from agroinfiltrated leaves, separated by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis and electrotransferred onto polyvinylidene diflouride (PVDF) membranes. Immunodetection was performed with anti-GFP antibodies. Red arrow indicates detection of G3890 fusioned protein only in nuclear protein extraction.

3.4. Discussion

Direct demonstration of effector entry into host cells remains to be a technical challenge for filamentous plant pathogens (Petre and Kamoun, 2014). However, the use of transient expression technique to study fungal effectors has proven to be effective to trap and detect plant protein partners (Petre *et al.*, 2015).

S. scitamineum G3890_chr10 putative effector seems to target exclusively nuclear compartment, including nucleolus. Results of RNAseq showed that plant responses to smut infection at 5 DAI (Chapter 3) include up-regulation of genes categorized in GO terms related to several transcriptional regulation mechanisms, suggesting a potential role of these effectors in activating/deactivating transcription of host genes which is worth to investigate.

Several effectors of fungi and oomycetes have been previously reported to localize into host nuclear compartments (Caillaud *et al.*, 2012; Schornack *et al.*, 2010; Vargas *et al.*, 2016), which can act as transcriptional regulators and involved in nucleocytoplasmic trafficking (Alfano, 2009). For example, in the oomycete *Hyaloperonospora arabidopsidis* the HaRxL44 effector interacts with MED19a, a subunit of the *Arabidopsis* Mediator complex, which directly interacts with RNA polymerase II and coordinate the action of many co-activators and co-repressors of gene

expression. Also, mediator complexes have been described to have function in the activation of signaling pathways, such as flowering, cell proliferation, production of small and long noncoding RNAs, regulators of organ size and phenylpropanoid homeostasis (Caillaud *et al.*, 2013b).

In the fungus *Melampsora larici-populina*, the MLP124017 effector is localized in the nucleus specifically associated with *N. benthamiana* and poplar TOPLESS and TOPLESS-related proteins (TPL/TPR), respectively. These transcriptional corepressors are involved in a wide range of processes including plant immune responses, such as repression of jasmonate signaling (Petre *et al.*, 2015).

Detection of two bands in western blot for the fused protein G3970_chr10, one with the expected size and another bigger, indicates the occurrence of post-translational modifications (PTM). PTMs of fungal effectors were previously described. For instance, N-glycosylation of LysM effector from *M. oryzae* is essential for its function in avoiding chitin perception by plant receptors (Chen *et al.*, 2014). N-glycosylation is also important in *U. maydis* effectors (Fernández-Álvarez *et al.*, 2013), where defective protein glycosylation mutants, such as the *O*-mannosyltransferase *pmt4* - essential for appressorium formation and penetration (Fernández-Alvarez *et al.*, 2009) - or the glucosidase II α-subunit *gas1* - crucial for growth inside the plant after appressorium penetration (Schirawski *et al.*, 2005) - exhibit severely compromised virulence. Analysis of G3970_chr10 protein sequence using *Prosite* to find potential PTM sites revealed the occurrence of cAMP- and cGMP-dependent protein kinase phosphorylation site, N-myristoylation site, Casein kinase II phosphorylation site, Protein kinase C phosphorylation site and N-glycosylation site. Since these modifications occurs in ER, G3970_chr10 localization surrounding the nucleus may be related to these PTMs.

3.5. Perspectives

Transient expression of plant pathogen effectors is one of the possible approaches to investigate plant compartment targeted, and a first step to determine host proteins able to interact. Hereafter, we aim to perform co-immunoprecipitation assays and determine the possible plant integrators of *S. scitamineum* effectors using spectrometric approaches, and thus associate with the possible roles of pathogen proteins in hijack host pathways. BiFc assays along with immunoblot analysis will be used to confirm protein-protein associations. To allow a more accurate identification of host target proteins we will test transient expression of pathogen effectors in sugarcane plants from tissue culture as a bait to perform immunoprecipitation.

Also, considering that transient expression is an artificial way to deliver fungal effector proteins in host cells, we aim to transform *S. scitamineum* with the respective fluorescent effectors to confirm their secretion and integration in host cells as well as the cell compartment targeted.

REFERENCES

Alfano, J.R. (2009) Roadmap for future research on plant pathogen effectors. *Mol. Plant Pathol.* **10**, 805–13.

Bouarab, K., Melton, R., Peart, J., Baulcombe, D. and Osbourn, A. (2002) A saponin-detoxifying enzyme mediates suppression of plant defences. *Nature* **418**, 889–892.

Burg, H.A. van den, Harrison, S.J., Joosten, M.H.A.J., Vervoort, J. and Wit, P.J.G.M. de (2006) *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. *Mol. Plant. Microbe. Interact.* **19**, 1420–30.

Caillaud, M.-C., Piquerez, S.J.M., Fabro, G., Steinbrenner, J., Ishaque, N., Beynon, J. and Jones, J.D.G. (2012) Subcellular localization of the Hpa RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. *Plant J.* **69**, 252–65.

Caillaud, M.M.-C., Asai, S., Rallapalli, G., et al. (2013) A downy mildew effector attenuates salicylic acid–triggered immunity in *Arabidopsis* by interacting with the host mediator complex He, S., ed. *PLoS Biol.* **11**, e1001732.

Caplan, J., Padmanabhan, M. and Dinesh-Kumar, S.P. (2008) Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. *Cell Host Microbe* **3**, 126–35.

Chen, X.-L., Shi, T., Yang, J., et al. (2014) N-glycosylation of effector proteins by an α-1,3-mannosyltransferase is required for the rice blast fungus to evade host innate immunity. *Plant Cell* **26**, 1360–76.

Djamei, A., Schipper, K., Rabe, F., et al. (2011) Metabolic priming by a secreted fungal effector. *Nature* **478**, 395–398.

Esse, H.P. van, Van't Klooster, J.W., Bolton, M.D., Yadeta, K.A., Baarlen, P. van, Boeren, S., Vervoort, J., Wit, P.J.G.M. de and Thomma, B.P.H.J. (2008) The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *Plant Cell* **20**, 1948–63.

Fernández-Alvarez, A., Elías-Villalobos, A. and Ibeas, J.I. (2009) The O-mannosyltransferase PMT4 is essential for normal appressorium formation and penetration in *Ustilago maydis*. *Plant Cell* **21**, 3397–412.

Fernández-Álvarez, A., Elías-Villalobos, A., Jiménez-Martín, A., Marín-Menguiano, M. and Ibeas, J.I. (2013) Endoplasmic reticulum glucosidases and protein quality control factors cooperate to establish biotrophy in *Ustilago maydis*. *Plant Cell* **25**, 4676–90.

Gan, P.H.P., Rafiqi, M., Ellis, J.G., Jones, D.A., Hardham, A.R. and Dodds, P.N. (2010) Lipid binding activities of flax rust AvrM and AvrL567 effectors. *Plant Signal. Behav.* **5**, 1272–5.

Giraldo, M.C., Dagdas, Y.F., Gupta, Y.K., et al. (2013) Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nat. Commun.* **4**, 617–621.

Hemetsberger, C., Herrberger, C., Zechmann, B., et al. (2012) The Ustilago maydis effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity Xu, J.-R., ed. *PLoS Pathog.* **8**, e1002684.

Hogenhout, S.A., Hoorn, R.A.L. Van der, Terauchi, R. and Kamoun, S. (2009) Emerging concepts in efector biology of plant-associated organisms. *Mol. Plant-Microbe Interact.* **22**, 115–122.

Houterman, P.M., Cornelissen, B.J.C. and Rep, M. (2008) Suppression of plant resistance gene-based immunity by a fungal effector Cormack, B.P., ed. *PLoS Pathog.* **4**, e1000061.

Jonge, R. De, Bolton, M.D. and Thomma, B.P. (2011) How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. *Curr. Opin. Plant Biol.* **14**, 400–406.

Jonge, R. de, Peter van Esse, H., Kombrink, A., et al. (2010) Conserved fungal LysM effector Ecp6 prevents

chitin-triggered immunity in plants. *Science (80-.).* **329**, 953–955.

Jonge, R. de and Thomma, B.P.H.J. (2009) Fungal LysM effectors: extinguishers of host immunity? *Trends Microbiol.* **17**, 151–7.

Kale, S.D., Gu, B., Capelluto, D.G.S., et al. (2010) External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell* **142**, 284–95.

Khang, C.H., Berruyer, R., Giraldo, M.C., Kankanala, P., Park, S.-Y., Czymmek, K., Kang, S. and Valent, B. (2010) Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. *Plant Cell* **22**, 1388–403.

Kmen, B. and Doehlemann, G. (2014) Inside plant: biotrophic strategies to modulate host immunity and metabolism. *Curr. Opin. Plant Biol.* **20**, 19–25.

Ökmen, B., Etalo, D.W., Joosten, M.H.A.J., Bouwmeester, H.J., Vos, R.C.H. de, Collemare, J. and Wit, P.J.G.M. de (2013) Detoxification of α-tomatine by *Cladosporium fulvum* is required for full virulence on tomato. *New Phytol.* **198**, 1203–1214.

Panstruga, R. and Dodds, P.N. (2009) Terrific protein traffic: the mystery of effector protein delivery by filamentous plant pathogens. *Science (80-.).* **324**.

Petre, B. and Kamoun, S. (2014) How do filamentous pathogens deliver effector proteins into plant cells? McDowell, J.M., ed. *PLoS Biol.* **12**, e1001801.

Petre, B., Saunders, D.G., Sklenar, J., Lorrain, C., Win, J., Duplessis, S. and Kamoun, S. (2015) Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *Mol. Plant. Microbe. Interact.* **28**, 689–700.

Presti, L. Lo, Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S. and Kahmann, R. (2015) Fungal effectors and plant susceptibility. *Annu. Rev. Plant Biol.* **66**, 513–545.

Schirawski, J., Böhnert, H.U., Steinberg, G., Snetselaar, K., Adamikowa, L. and Kahmann, R. (2005) Endoplasmic reticulum glucosidase II is required for pathogenicity of *Ustilago maydis*. *Plant Cell* **17**, 3532–43.

Schornack, S., Damme, M. van, Bozkurt, T.O., Cano, L.M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S. and Huitema, E. (2010) Ancient class of translocated oomycete effectors targets the host nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17421–6.

Sekhwal, M.K., Li, P., Lam, I., Wang, X., Cloutier, S. and You, F.M. (2015) Disease resistance gene analogs (RGAs) in plants. *Int. J. Mol. Sci.* **16**, 19248–19290.

Shabab, M., Shindo, T., Gu, C., et al. (2008) Fungal effector protein AVR2 targets diversifying defense-related cys proteases of tomato. *Plant Cell* **20**, 1169–83.

Skibbe, D.S., Doehlemann, G., Fernandes, J. and Walbot, V. (2010) Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. *Science (80-.).* **328**, 89–92.

Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., Hoorn, R.A.L. van der and Kamoun, S. (2009) Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1654–9.

Takahara, H., Hacquard, S., Kombrink, A., et al. (2016) *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. *New Phytol.* **211**, 1323–1337.

Tanaka, S., Brefort, T., Neidig, N., et al. (2014) A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. *Elife* **3**, e01355.

Taniguti, L.M., Schaker, P.D.C., Benevenuto, J., et al. (2015) Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with Sugarcane. *PLoS One* **10**, e0129318.

Vargas, W.A., Sanz-Martín, J.M., Rech, G.E., Armijos-Jaramillo, V.D., Rivera, L.P., Echeverria, M.M., Díaz-Mínguez, J.M., Thon, M.R. and Sukno, S.A. (2016) A fungal effector with host nuclear localization and DNA-binding properties is required for maize Anthracnose development. *Mol. Plant-Microbe Interact.* **29**, 83–95.

Vleeshouwers, V.G.A.A. and Oliver, R.P. (2014) Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Mol. Plant-Microbe Interact.* **27**, 196–206.

Wahl, R., Wippel, K., Goos, S., Kämper, J. and Sauer, N. (2010) A novel high-afinity sucrose transporter is required for virulence of the plant pathogen *Ustilago maydis* Heitman, J., ed. *PLoS Biol.* **8**, e1000303.

Win, J., Chaparro-Garcia, A., Belhaj, K., et al. (2012) Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 235–247.

4. TRANSCRIPTIONAL PROFILE OF SUGARCANE IN SMUT DISEASE: FROM EARLY INFECTION TO WHIP DEVELOPMENT

The results presented in this chapter were published in the article "as part of the article entitled "RNAseq Transcriptional Profiling following Whip Development in Sugarcane Smut Disease" DOI:10.1371/journal.pone.0162237 in September 1, 2016. No permission is required from the authors or the publishers to reuse or repurpose PLOS content provided the original article.

Abstract

Sugarcane smut disease is caused by the basidiomycete fungus *Sporisorium scitamineum*, which establishes a biotrophic interaction. The disease is characterized by the development of a whip-like structure from the primary meristems, where billions of teliospores are produced. Other smut symptoms include tillering, low sucrose and high fiber contents, reducing cane productivity. We suggested the biological events contributing to disease symptoms in a smut intermediate resistant sugarcane genotype by examining the transcriptional profiles (RNAseq) shortly after inoculating the plants and immediately after whip emission. The overall picture of disease progression suggests a premature transcriptional reprogramming of the shoot meristem functions continuing until the emergence of the whip. The guidance of this altered pattern is potentially related primarily to auxin mobilization in addition to the involvement of other hormonal imbalances. The consequences associated with whip emission are modulation of typical meristematic functions toward reproductive organ differentiation requiring strong changes in carbon partitioning and energy production. These changes include overexpression of genes coding for invertases and trehalose-6P synthase, and other enzymes from key metabolic pathways, such as that of the lignin biosynthesis. This is the first report of changes in transcriptional profiles following whip development providing a hypothetical model and candidate genes to further study sugarcane smut disease progression.

4.1. Introduction

Sugarcane (*Saccharum* spp.) is the fifth most important crop in the world (FAOSTAT, 2013). Besides being a source of sugar for food, the crop has the potential to generate clean and renewable products such as biofuels, bioplastics, bio-hydrocarbons, and bioelectricity. Due to its agronomic attributes such as high yield and survival under adverse conditions (Waclawovsky *et al.*, 2010), sugarcane is found in more than 100 tropical and subtropical countries (FAOSTAT, 2013). Nonetheless, the crop hosts several pathogens, including the fungus *Sporisorium scitamineum*, the causal agent of sugarcane smut disease (Sundar et al., 2012) (Fig 1). Sugarcane smut is mainly characterized by the development of a long whip-like structure consisting of plant and fungal tissues where billions of teliospores are produced. The name 'smut' derives from the black powdery mass of teliospores released by these structures that resemble soot. The whips originate in the primary meristems of the apex and lateral buds of infected stalks, and they are initially covered with a thin silvery membranous sheath (Sundar *et al.*, 2012), which detaches after teliospores mature and are ready to disperse. In the more susceptible varieties, whips can be detected as early as 2 to 4 months of age, with peak whip growth occurring in the 6th or 7th month (Legaz *et al.*, 1998). Smut is mainly transmitted by wind-borne teliospores infecting the standing canes, but also by teliospores in the soil that infect the planted setts. The germination of the teliospores leads to meiosis, which produces haploid sporidia. Mating compatible sporidial cells produce infective hyphae through hyphal anastomosis, which initiates plant colonization (Bakkeren *et al.*, 2008).

The disease limits the crop yield and properties of sugarcane products, causing losses in cane tonnage and juice quality. Other disease symptoms include tillering and low sucrose and increased fiber contents (Sundar *et al.*, 2012). Like most agronomic traits, smut resistance is a quantitative character (Chao *et al.*, 1990) that is difficult to genetically and functionally characterize. Moreover, modern varieties of sugarcane (2n = 100–130) have a complex genomic structure that derives from a highly polyploid and aneuploid interspecific hybridization (D'Hont *et al.*, 1996; de Setta *et al.*, 2014), hindering the understanding of the quantitative traits and mapping their loci (Garcia *et al.*, 2013; Palhares *et al.*, 2012). Efforts to elucidate the molecular basis of sugarcane smut resistance have been made since James (1973) proposed the existence of a chemical resistance mechanism. Lloyd and Pillay (1980) identified some flavonoids, which are teliospore-germination inhibitors, and subsequently a correlation between the resistance rating and the concentration of glycosidic substances was established (Lloyd, 1983). Later studies reported changes in the patterns of free polyamines and their conjugation in both susceptible and resistant sugarcane varieties infected by *S. scitamineum* (Legaz *et al.*, 1998; Piñon *et al.*, 1999).

Changes in the sugarcane gene expression profile induced by the fungus have been identified by several authors using techniques such as suppression-subtractive hybridization-based sequencing and the differential display of complementary DNA-amplified fragment-length polymorphisms (Borrás-Hidalgo *et al.*, 2005; LaO *et al.*, 2008; Rutherford, 2001; You-Xiong *et al.*, 2011).

Despite these attainments, more detailed studies are needed to precisely define the changes in the entire sugarcane gene repertoire when challenged with the pathogen, both at different stages of fungal development and in different host tissues. Messenger RNA sequencing (RNAseq) technology has the potential to explore the complete set of gene expression programs to a high level of accuracy and depth, providing further insights into the plant-pathogen interactions (Westermann *et al.*, 2012). This method has been applied to several mixed-model systems of plant-fungus interactions (Kawahara *et al.*, 2012; Lowe *et al.*, 2014; Tremblay *et al.*, 2011; Yazawa *et al.*, 2013; You-Xiong *et al.*, 2011; Zhu *et al.*, 2013), and more recently, to elucidating the early stages of the sugarcane-smut pathosystem (Que, Su, *et al.*, 2014; Taniguti *et al.*, 2015; Wu *et al.*, 2013). Continuing the study of this pathosystem, we used RNAseq technology to perform a comparative analysis of infected sugarcane tissues of a smut intermediate resistant genotype at two time points: shortly after inoculation and later, when the whips appeared and disease symptoms were evident (Figure 13). Besides confirming existing data (Que, Su, *et al.*, 2014), this work address the molecular events following whip emission. The most relevant conclusions are: 1) the association with transcriptional reprogramming of shoot apical functions probably by restraining floral development; 2) transcriptional changes in carbon partitioning, mostly pronounced towards hexoses and lignin; and 3) the auxin seems the relevant hormone related to whip emission as well as the response associated with oxidative stress.

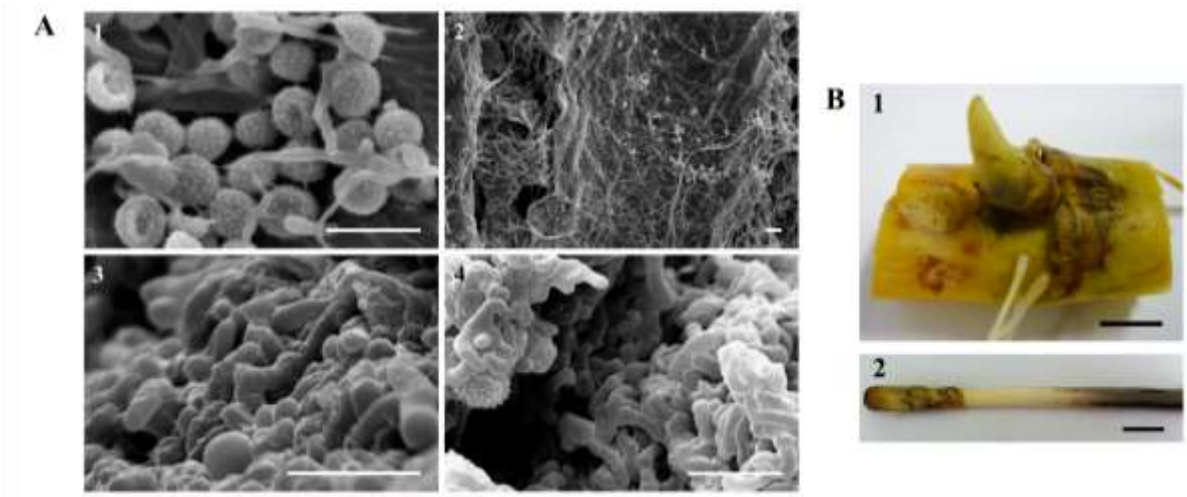


Figure 13. Sugarcane smut. (A) Scan electron microscopy of *S. scitamineum* hyphal growth in sugarcane bud at 5 DAI (1, 2); and fungal sporogenesis and teliospores maturation in the base of the sugarcane whip 200 DAI (3; 4). Bar = 10 μm. (B) (1) Sugarcane bud 5 DAI; and (2) base of the whip 200 DAI. Bar = 1 cm.

4.2. Material and Methods

4.2.1. Ethics Statement

S. scitamineum SSC39 teliospores were collected as described by Taniguti *et al.* (2015). RNAseq experiments were performed using the smut intermediate resistant Brazilian commercial variety of sugarcane, 'RB925345'. The healthy buds used to conduct the experiments were obtained from the IAC (Instituto Agronômico de Campinas, Centro de Cana), Ribeirão Preto, São Paulo, Brazil. No special permits were necessary for the teliospores or cane collection.

4.2.2. Experimental design

S. scitamineum SSC39 teliospores were checked for viability and inoculated as previously described by artificial wounding method (Taniguti *et al.*, 2015). The initial sugarcane response was analyzed based on pools of 10 breaking buds collected 5 DAI (days after inoculation). The late response was evaluated using culms after the whips emerged 200 DAI. Sampling was at the base of the whips, up to 2 cm below the culm. This is a region of intensive sugarcane cell division and fungal sporogenesis. The infected plants were compared to control (mock-inoculated) plants of the same age. Three biological replicates were included for each inoculated and control treatment using a completely randomized design maintained on greenhouse benches (Figure 14). PCR amplicon containing the rDNA internal transcribed spacer region (ITS1, 5.8S and ITS2) of *S. scitamineum* generated with primers: Hs 5' -AACACGGTTGCATCGGTTGGGTC- 3' and Ha 5' -GCTTCTTGCTCATCCTCACCACCAA- 3') according to Bueno (2010) was used to confirm infection 5 DAI.

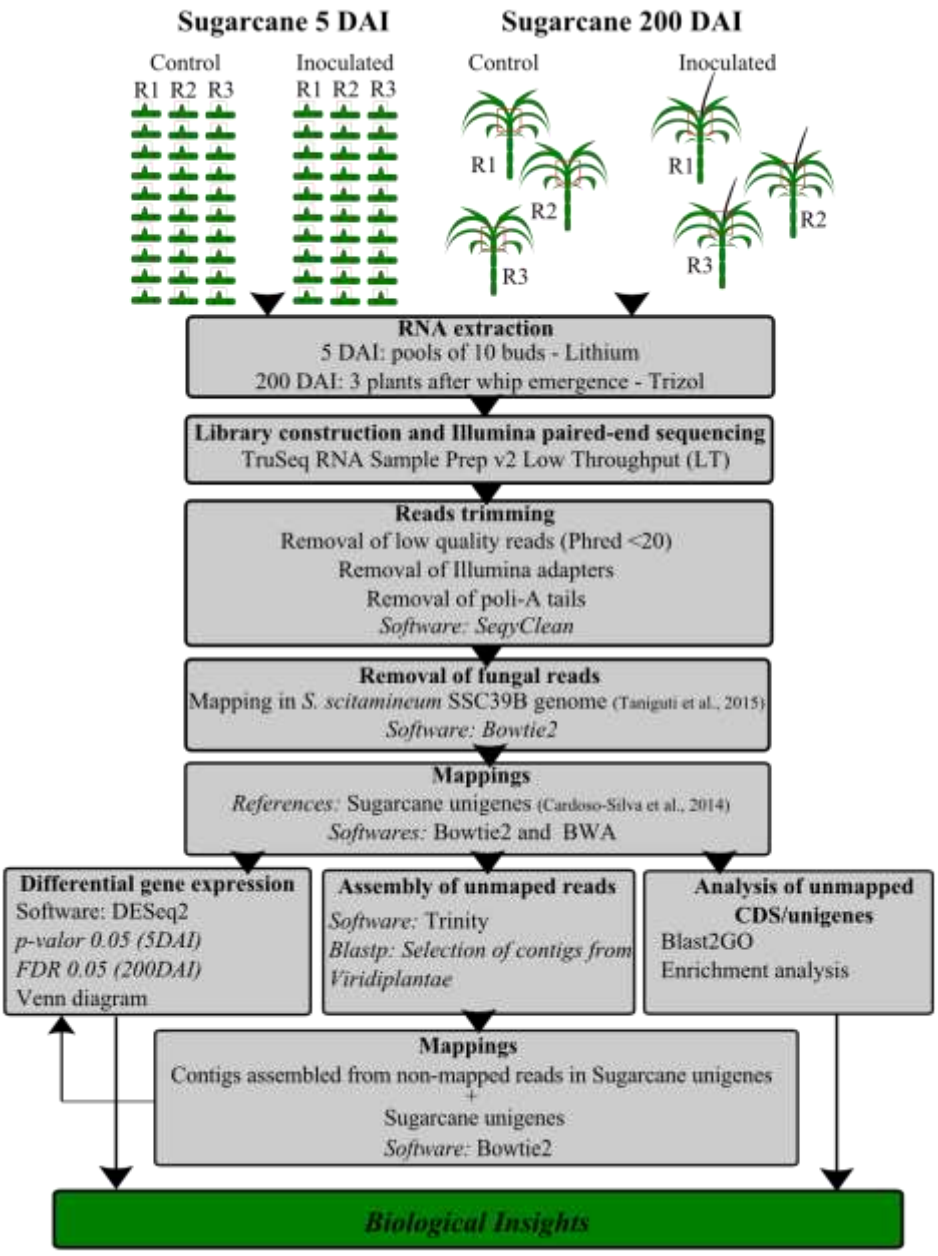


Figure 14. Experimental design to investigate transcriptional changes of sugarcane plants of the intermediate resistant variety "RB925345" in response to *S. scitamineum* development. The time points analyzed were: 1) 5 DAI (days after inoculation); and 2) 200 DAI after whip emission. Single budded sets of seven month-old plants were surface disinfected, heat treated (52°C for 30 min in water bath, 1 kg of buds/6L of water) and incubated for 16 h at 28°C. Artificial inoculation was performed using the paste method in previously needle damaged buds to overcome mechanical/pre-formed resistance. Mock inoculated plants were used as control. A greenhouse experiment was conducted in completely randomized design with two treatments (5 DAI and 200 DAI) and three replications: 1) three pools of 10 breaking buds were used to determinate transcriptional changes five days post-mock or inoculation with *S. scitamineum;* and 2) three diseased plants after whip emission and three healthy plants of the same age were used to determinate transcriptional changes 200 DAI. RNA extraction methods used were: 1) 5 DAI, lithium-based protocol (Gasic et al., 2004); and 2) 200 DAI, TRIzol® (Life Technologies #15596-018) according to manufacturer's instructions.

4.2.3. RNA extraction, libraries, and sequencing

Total RNA was extracted from the samples using distinct methods for each plant developmental stage as described by Taniguti *et al.* (2015) (Fig 14). The quality of the total RNA was verified using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA), and the libraries

were constructed using the TruSeq RNA Sample Prep v2 Low Throughput (LT) kit as described in the manufacturer's instructions (Illumina, San Diego, CA). The libraries were paired-end sequenced using the Illumina system (HiScanSQ).

4.2.4. Pre-processing and mapping the Illumina reads

The Illumina reads were treated as previously described (Taniguti *et al.*, 2015) (Figure 14). Two reference sequences were used for mapping the RNAseq data: the complete genome sequence of *S. scitamineum* (Taniguti *et al.*, 2015); and a set of unigenes produced by the assemblage of RNAseq data from six sugarcane cultivars (Cardoso-Silva *et al.*, 2014). The software packages used for the mapping were Bowtie2 V2.1.0 (30) and BWA (Li and Durbin, 2009). Bowtie2 was used with the default parameters in sensitive mode (-D 15; -R 2; -L 22; -i S, 1, 1.15), while the BWA alignments were obtained using the default parameters (-n 0.04; -k 2; -O 11). The RNAseq reads that showed no similarities to the sugarcane unigenes using the above parameters were assembled using Trinity (Grabherr *et al.*, 2011). Clusters identified by the prefix "gg" were then selected by comparison to the Viridiplantae sequences of UniProtKB (Uniprot Consortium, 2015).

4.2.5. Sugarcane gene expression analysis

The differentially expressed genes (DEGs) were identified using the DESeq2 package (Love *et al.*, 2014). For the 5-DAI data, DEGs were considered statistically significant if they had a p-value less than 0.05 when compared to the control buds. The multiple-test correction proposed by Benjamini and Hochberg (Benjamini and Hochberg, 1995) was used for the 200-DAI data by applying a FDR (False discovery rate) to generate a set of DEGs with the same significance level (<0.05). The DrawVenn webtool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to produce Venn diagrams from the sets of DEGs obtained from the BWA/DESeq2 or Bowtie2/DESeq2 analyses and different reference sets (Figure 14).

4.2.6. Annotation and Gene ontology analysis

The BLAST2GO tool V2.7.2 (Conesa and Götz, 2008) was used with the default parameters to assign GO (Gene Ontology) terms to the DEGs. Metabolic pathways analysis was performed based on the KEGG database (Kanehisa and Goto, 2000). The GO enrichment analysis was conducted with the BLAST2GO tool using the two-sided Fisher's Exact Test with

the p-value set at ≤ 0.05. The GenBank (Benson *et al.*, 2000) and UniProt (Unirprot Consortium, 2015) databases and the InterProScan (Jones *et al.*, 2014), SignalP (Petersen *et al.*, 2011), TMHMM (Krogh *et al.*, 2001), ScanProsite, and MyDomains (de Castro *et al.*, 2006) tools were used to predict function and features of protein sequences.

4.2.7. Quantitative PCR (qPCR) expression analysis

Quantitative PCR analysis was used to confirm the gene expression profiling data obtained from the RNAseq. Transcripts encoding: invertase, auxin transporter, trehalose-6P synthase, pyruvate decarboxylase, aldolase, S-adenosylmethionine synthetase (SAM), peroxidase and the longifolia-like protein (LGN) were selected for the reverse transcription-qPCR (RT-qPCR) reactions (Table 6). The primers were manually designed and the quality verified using Gene Runner (http://www.generunner.net/) and NetPrimer (http://www.premierbiosoft.com/netprimer/). All RT-qPCRs were performed in the 7500 Fast Real-Time PCR System (Applied Biosystems, Waltham, MA) using the GoTaq® One-Step RT-qPCR System Kit (Promega, Madison, WI). A reaction mixture containing 50 ng of RNA, 6.5 µL of GoTaq® qPCRMaster Mix, 0.2 µM of each primer, 0.25 µL of GoScript™ RT Mix, and nuclease-free water to a final volume of 12.5 µL was used for the three biological replicates and two technical replicates. Cycling conditions were as follows: 37 °C for 15 min; 95 °C for 10 min; 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s. Primer specificity was confirmed by obtaining the dissociation curve for each reaction. Sugarcane housekeeping genes encoding for polyubiquitin (Papini-Terzi, 2005) and GAPDH (d-glyceraldehyde-3-phosphate dehydrogenase; (Iskandar *et al.*, 2004) were used to normalize the expression signals. The PCR efficiencies and Cq values were obtained using the LinReg PCR program (Ramakers *et al.*, 2003). Relative changes in the gene expression ratios were calculated with REST software (Pfaffl *et al.*, 2002). Control samples (mock-inoculated plants) were used as calibrators. The Student *t*-test was used to estimate significant changes in the relative expression levels ($p < 0.05$).

Table 6. Primers used to RTqPCR validation.

Product	Transcript	Primer sequence
GAPDH	Iskandar et al., 2004	5' CACGGCCACTGGAAGCA 3'
		5' TCCTCAGGGTTCCTGATGCC 3'
Polyubiquitin	Papinni-Terzi et al., 2005	5' CCGGTCCTTTAAACCAACTCAGT 5'
		3' CCCTCTGGTGTACCTCCATTTG 3'
Invertase	comp201528_c0_seq1	5' GGAGGACGAGACCACACTC 3'
		5' CGTTGTTGAAGAGGAACAC 3'
Auxin transporter	comp205699_c0_seq1	5' GCTCCTGACTCTGCCGTAC 3'
		5' TGACGTGGTTCTTGAAGCTG 3'
Trehalose-6P synthase	comp204716_c0_seq1	5' TGCCGATCTGATTGGGTTCC 3'
		5' GCTCCAAGTGAACTCCCACA 3'
Pyruvate decarboxylase	comp200606_c0_seq1	5' CGACGGACCATACAACGTCA 3'
		5' GCACAGGCAGTCCTTCTTCT 3'
Aldolase	comp196354_c1_seq1	5' CTGAGGTGATTGCTGCGTAC 3'
		5' GACAGGGACCACGGCTTCT 3'
S-adenosylmethionine (SAM)	comp194455_c0_seq1	5' GTTGGTCTTGGTGCAGGTCT 3'
		5' GAGAAGATGGCCGCAGTTGA 3'
Peroxidase	comp187834_c0_seq1	5' GTCTCGTCGGTGTAGAGCAC 3'
		5' TACAGCTACAGCAGCAGCAC 3'
Longifolia like	comp200950_c0_seq1	5' AGAAGCATGGGGTTTCACTG 3'
		5' GGCTTGATGAGCTTGTAGGC 3'

4.3. results and discussion

4.3.1. General analysis

In this study, the smut intermediate resistant variety 'RB925345' developed whips and other disease symptoms beginning 127 days after inoculation (DAI). However, the plants were sampled 200 DAI, because this was the time when whips were detected in all three replicate plants used in the experimental design (Figure 14). Out of the total number of inoculated plants, 48 (53%) developed whip over the timeline of the experiment (334 days). (Figure 15A).

Samples of 5 DAI were used to amplify the 509-bp sequence that corresponds to the 5.8S ribosomal RNA gene and flanking internal transcribed spacers 1 and 2 in *S. scitamineum* (Bueno, 2010) confirming the fungal infection (Figure 15B). This was particularly necessary to confirm infection 5 DAI, since buds were collected with no smut disease symptoms.

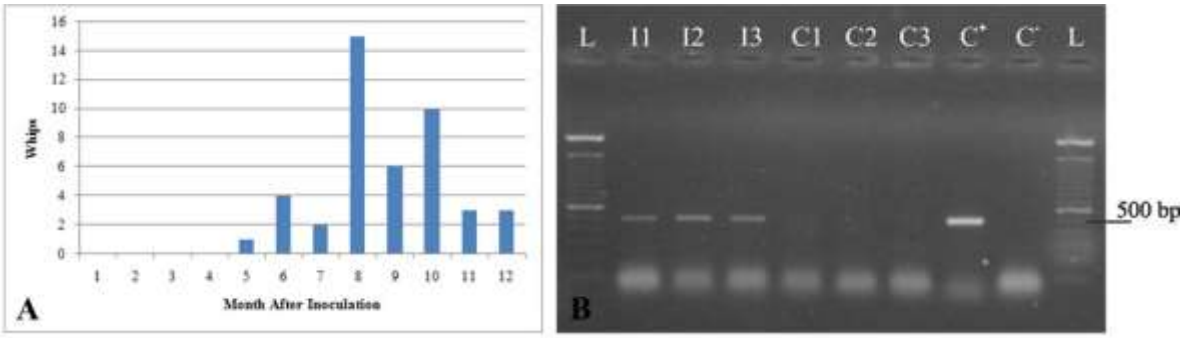


Figure 15. A) Number of whips developed each month after smut inoculation in the intermediate resistant genotype RB925345. B) Amplicons of primers Hs and Ha using total DNA of buds collected 5 DAI. L: ladder, I: infected, C: control.

A total of 225.2 million paired-end sequences (PEs) of ~100 bp (~22.5 Gbp) were obtained for the 12 RNAseq libraries (~18 million reads per library). Including the corresponding control libraries, 111,926,958 (49.7%) PEs were from the 5-DAI collection and 113,269,226 (50.3%) PEs were from 200-DAI (Table 7). Fungal sequencing reads were screened out after mapping them to the whole *S. scitamineum* SSC39B genome (Taniguti *et al.*, 2015), leading to the removal of approximately 20% (2% 5 DAI; 18% 200 DAI) of the PEs.

Table 7. Total of RNAseq data obtained.

RNAseq library	Raw data	Number of high quality reads
5DAI_I1	16,266,804	14,884,654
5DAI_I2	21,017,222	19,101,404
5DAI_I3	19,075,478	17,390,580
5DAI_C1	17,166,848	15,676,736
5DAI_C2	18,564,444	16,912,200
5DAI_C3	19,836,162	18,132,756
5 DAI_*total* **(%)**	111,926,958 (**49.75%**)	102,098,330 (**91.2%**)
200 DAI_I1	17,705,938	13,432,560
200 DAI_I2	21,691,820	14,911,384
200 DAI_I3	17,617,212	12,553,870
200 DAI_C1	19,382,530	15,758,474
200 DAI_C2	18,126,950	14,750,944
200 DAI_C3	18,744,776	15,287,676
200 DAI_*total* (%)	113,269,226 (**50.3%**)	86,694,908 (**76.5%**)
Total (%)	225,196,184 **(100%)**	18,8793,238 (**74%**)

4.3.2. Count-based differential expression analysis of the RNAseq data

A set of previously obtained sugarcane transcripts was used to describe the biological events underlying the interaction with *S. scitamineum*. The reference set of sugarcane unigenes consisted of 72,268 sequences obtained from a de novo RNAseq assembly and a transcriptome annotation for six cultivars collected in various sugarcane crop fields, including the 'RB925345' variety (Cardoso-Silva *et al.*, 2014). This set of unigenes was used to allow cross-comparisons between sugarcane sequencing data. To define the best alignment of the RNAseq to the unigenes set we performed two analyses. Using both Burrows-Wheeler Aligner (BWA) (Li and Durbin,

2009) and Bowtie2 (Langmead and Salzberg, 2012) softwares, approximately 73% of the high quality sequence reads were aligned to 67% of the unigenes (Table 8).

Table 8. Mapping results of BWA and Bowtie2 softwares using sugarcane unigenes as reference.

		Sugarcane Unigenes			
		Reads Mapped		Unigenes Mapped	
RNAseq library	**Number of high**	**BWA**	**Bowtie2**	**BWA**	**Bowtie2**
5DAI_I1	14,884,654	10,733,379	11,060,027	46,227	46,513
5DAI_I2	19,101,404	13,837,994	14,225,046	46,751	48,018
5DAI_I3	17,390,580	12,524,618	12,845,628	46,981	48,306
5DAI_C1	15,676,736	11,342,026	11,685,978	46,099	46,641
5DAI_C2	16,912,200	12,549,339	12,690,667	47,714	46,583
5DAI_C3	18,132,756	12,760,642	13,172,172	47,952	47,397
5 DAI_*total* **(%)**	102,098,330	73,747,998 **(72%)**	75,679,518 **(74%)**	–	–
200 DAI_I1	13,432,560	10,187,611	10,380,877	38,922	38,934
200 DAI_I2	14,911,384	10,217,800	10,483,746	44,559	44,868
200 DAI_I3	12,553,870	8,737,145	8,886,936	41,424	41,466
200 DAI_C1	15,758,474	11,108,875	11,364,603	47,418	47,765
200 DAI_C2	14,750,944	10,416,216	10,655,891	45,756	45,903
200 DAI_C3	15,287,676	10,850,245	11,095,305	47,202	47,480
200 DAI_*total* **(%)**	86,694,908	61,517,892 **(70%)**	62,867,358 **(72%)**	–	–
Total (%)	188,793,238	135265890 **(71.6%)**	138,546,876 (73**%**)	–	–

The remaining subset of reads (15,000,000 PEs), those that showed no similarity to the sugarcane unigenes, were clustered using Trinity v2.0.6 (Grabherr *et al.*, 2011). A total of 25,794 contigs with more than 500 bp were assembled, and 16,219 were defined as 'RB925345' transcripts based on the presence of orthologs in the Viridiplantae section of the UniProt database (UniProt release 2015_03). These transcripts were identified by the prefix "gg" and probably include those most related to the pathogen infection. This new set of transcripts (gg) was combined with the sugarcane unigenes (29) (88,487 transcripts) to define the final set of DEGs (Figure 16). Our goal was to understand differential gene expression starting shortly after inoculation (5-DAI) and continuing through fungal sporogenesis subsequently whip emission (200-DAI). DEGs were defined for both 5-DAI and 200-DAI samples and the intersection between the two transcript groups were used for the annotation processes (Figure 16).

It was not possible to detect DEGs using FDR in 5-DAI samples. However, we pursued the analysis further cautiously using p-values set at less than 0.05 and 0.01. The subtle sugarcane response shortly after inoculation may be due to low percentage of the fungus in buds and/or delayed plant response due to the susceptibility of the variety used in the experiment (Que, Su, *et al.*, 2014). Only 2% of the total reads were detected as genes expressed by the fungus, which represents around 67% of the *S. scitamineum* complete set of genes (Taniguti *et al.*, 2015).

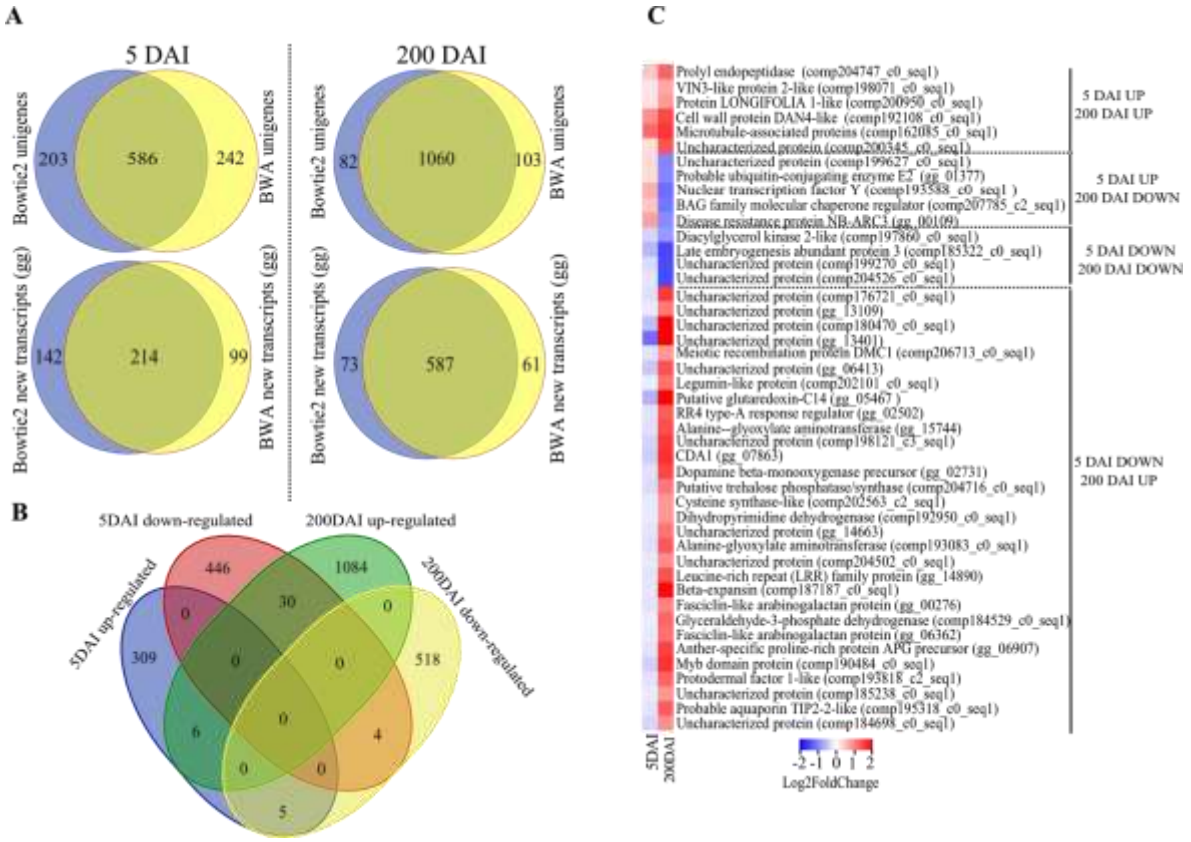


Figure 16. Differential expression gene analysis of 5-DAI and 200-DAI samples of the smut infected plants. General results of differential expression analysis in early and late time points of the smut compatible interaction. A) Venn diagrams show number of genes detected as differentially expressed by DESeq2 package (Love et al., 2014) using two mapping softwares (Bowtie2 and BWA) and two sets of references: sugarcane unigenes (Cardoso-Silva *et al.*, 2014) and the novel sugarcane transcripts assembled from sugarcane unigenes unmapped reads using Trinity (Grabherr *et al.*, 2011). Intersections of the identified DEGs from the two mapping results were considered as the final set of DGEs for the following analysis. **B)** Venn diagram show the final set of DEGs up- and down-regulated of DEGs from 5 and 200 DAI samples. **C)** Expression profile of 45 DEGs shared between 5 and 200 DAI (Inoculated/Control). Heatmap constructed using R software with gplots package.

4.3.3. Enrichment analysis of GO terms

The molecular events underlying sugarcane response during infection were suggested initially based on GO terms assignment and GO enrichment analysis of DEGs. The enrichment analysis of 5-DAI DEGs showed as expected two contrasting molecular responses as previously described (Que, Su, *et al.*, 2014). Genes involved in general plant immunity were down-regulated, while those for epigenetic mechanisms were up-regulated (Figure 17). In addition, this same enrichment test included terms related to shoot apical activities with the identification of three GO terms: Regionalization; Organ boundary specification; and Specification of floral organ identity. These terms suggest that the plant meristem functions are prematurely modulated by the presence of the pathogen. Some genes related to this same functional group were also enriched after whip emission. The gene regulatory network for shoot apical functions known in plant

models is responsible for the differentiation of cells and organs (leaves and inflorescences). In corn, smut fungi are known to prevent or to modify floral organ differentiation inducing tumor-like galls (Gao *et al.*, 2013; Ghareeb *et al.*, 2011). We suggest that a similar modulation occurs in susceptible genotypes of sugarcane infected with *S. scitamineum* very shortly after colonization. However, even though GO terms enrichment test lean towards that, this hypothesis needs to be further investigated since 5-DAI DEGs were not supported by FDR. A time-course experiment using sugarcane varieties showing various levels of smut-related response, for instance, should be conducted to determinate the expression profile of genes associated with meristematic functions identified here.

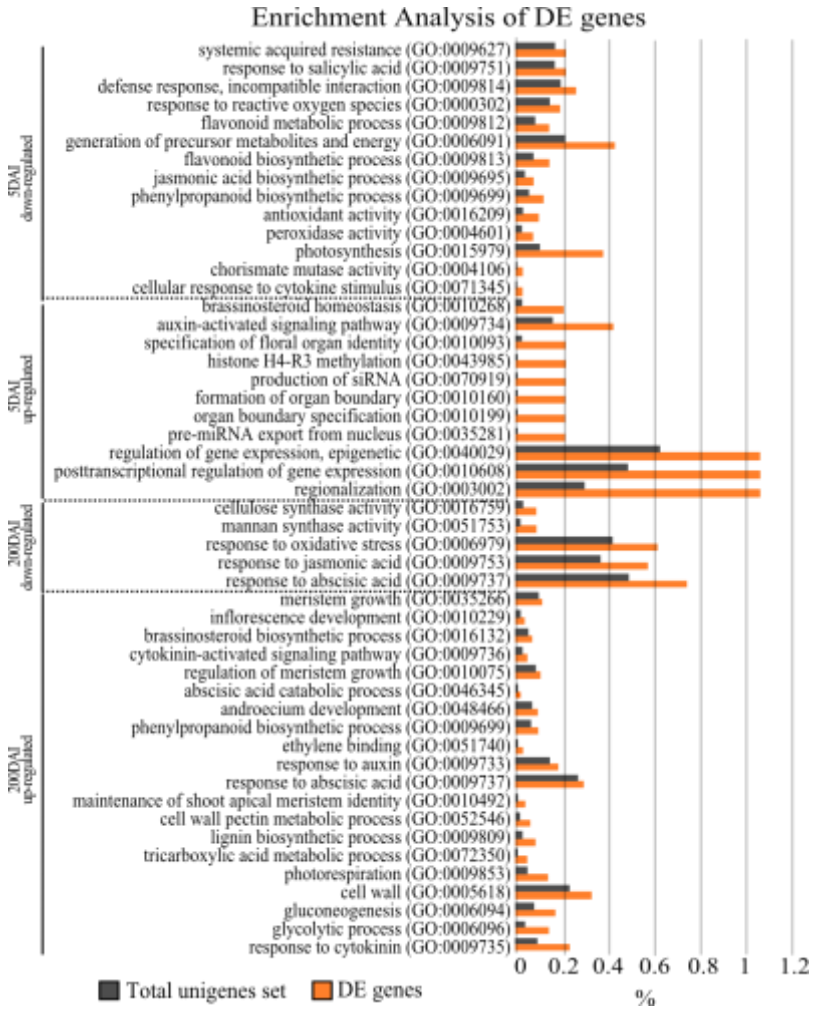


Figure 17. Enrichment analysis of GO terms. DEGs were submitted to enrichment analysis in BLAST2GO software, p-value ≤ 0.05 was used as cut-off parameter. Grey bars represent the percentage of genes related to each selected GO term in the total set of sugarcane unigenes. Red bars represent the percentage of genes related to each selected GO term in the set of DEGs. The complete list of enriched GO terms in each set of DEGs can be found in Schaker et al., 2016 (File S6).

Meristem-related functions were remarkably affected considering the enrichment test of 200-DAI sample DEGs. A number of genes identified allowed us to propose a putative model

(Figure 18) for the transition of the normal meristematic functions controlled by the interaction of auxin and cytokinin hormones (Su *et al.*, 2011) to the development of the whip. For instance, a *longifolia-like* gene (LNG, *comp200950_c0_seq1*) is up regulated at this time point. Mutants overexpressing this same gene in *A. thaliana* have long petioles, narrow but extremely long leaf blades with serrated margins, elongated floral organs, and elongated siliques as a result of polar-cell elongation (Lee *et al.*, 2006). This description resembles the whip development in sugarcane. LNG-like gene was assayed by qPCR confirming the RNAseq data (Figure 19). Transcripts encoding homologs of VIN3 (*vernalization insensitive 3 protein*) were also up-regulated after whip emission. VIN3-like proteins are involved in both the vernalization and photoperiod pathways by regulating expression of floral repressors FLOWERING LOCUS C (FLC) and FLOWERING LOCUS M (FLM). In *A. thaliana*, VIN3-LIKE protein epigenetically represses a member of the FLC family, enabling flowering in non-inductive photoperiods (Kim and Sung, 2014). Additional transcripts members of three gene classes responsible for floral development (A, B and C) were identified. The ABCDE model proposes that a certain combination of MADS proteins activates different groups of genes related to flowering (Honma and Goto, 2001; Murai, 2013). A MADS-box TF homologous to AP1 (APETALA1; class A) was detected highly expressed (*comp207551_c1_seq1*; $\log_2$ fold-change = 9.4). This gene is essential in *A. thaliana* for the transition from inflorescence meristem to a floral meristem (Kaufmann *et al.*, 2010). This same MADS-box TF, along with the product of FLOWERING LOCUS T (FT; *gg_11173*), which is also highly expressed in infected sugarcane plants ($\log_2$ fold-change = 2.74), promotes the transition from vegetative to reproductive growth. In *A. thaliana*, FT encodes a small peptide recognized as the major component of florigen that induces the expression of other floral genes such as AP1 (Eckardt, 2010). Three other MADS-box TFs were up-regulated after whip emission, encoding homologs of APETALA3 (AP3, *gg_00300*), the class B gene AGAMOUS (AG*, comp204141_c1_seq1*), the class C gene APETALA1 (AP1, *gg_05696*) and COL6 (C2C2-CO-like transcription factor; *comp194394_c0_seq1*). COL6 belongs to the CONSTANS family and encodes a putative zinc finger TF promoting the induction of flowering in *A. thaliana* during long photoperiods (Lagercrantz and Axelsson, 2000) through the activation of floral meristem-identity genes such as LEAFY (Simon *et al.*, 1996). Regulatory switches coordinating these developmental changes have been extensively studied in *A. thaliana* (Kaufmann *et al.*, 2010); they are very precise and could vary in sugarcane, but the enrichment of genes related to the transition in meristem functions led us to associate these events with the plant-pathogen interaction mode (Kaufmann *et al.*, 2010). It seems reasonable to assume that a combination of MADS-box TFs that are up-

regulated in smut-infected plants may coordinate the gene expression related to whip development as an alternative route instead of the normal flowering program.

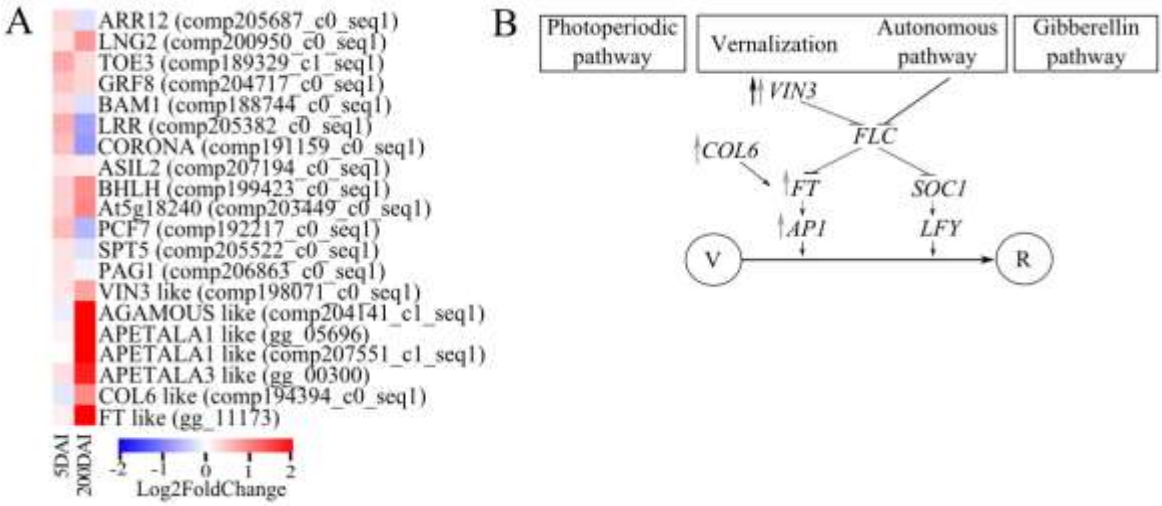


Figure 18. Sugarcane DEGs related to meristem functions. (A) Expression profile of genes related to meristem functions represented as values of log2 fold change (Inoculated/Control). Heatmap was constructed in R software package. Blue squares represent down-regulated genes, red squares represent up-regulated ones. (**B)** Model of probable events related to whip development in sugarcane. Increase in VIN3 expression early in infected plants may release FT expression, which in turn positively regulates Apetala-1 (AP1) expression, turning the vegetative growth program to reproductive, via the autonomous/vernalization pathway. Black arrow represents up-regulation at 5 DAI, grey arrows represent up-regulation at 200 DAI.

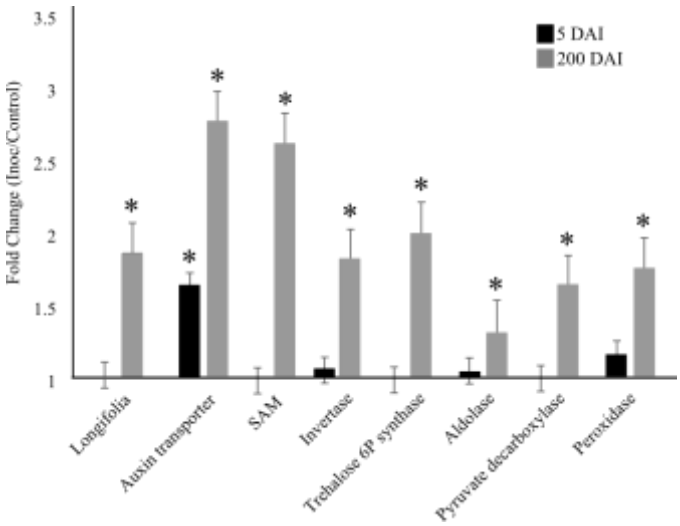


Figure 19. RT-qPCR validation. Sugarcane unigenes selected for RT-qPCR analysis of 5-DAI and 200-DAI samples: longifolia-like protein (*comp200950_c0_seq1*); auxin transporter (*comp205699_c0_seq1*); SAM (*comp194455_c0_seq1*); invertase (*comp201528_c0_seq1*); trehalose 6P synthase (*comp204716_c0_seq1*); aldolase (*comp196354_c1_seq1*); pyruvate decarboxylase (*comp200606_c0_seq1*) and peroxidase (*comp187834_c0_seq1*). Reactions were performed using one-step GoTaq® One-Step RT-qPCR System Kit (Promega) in a 7500 Fast Real-Time PCR System (Applied Biosystems). Statistical analysis was performed in REST® software. "*" indicates genes differentially expressed in RT-qPCR reactions (p-value < 0.05).

Given the interaction of *S. reilianum* with maize (Ghareeb *et al.*, 2011) and the data we presented here, both *Sporisorium* species may share a common trend of modifying meristem

identity. In maize, phyllody and tumor formation result from alterations in the floral developmental program at both the apex and axillary meristems (Ghareeb *et al.*, 2011). In sugarcane, the whip may result from releasing the transition from vegetative to reproductive/flowering, potentially via the autonomous/vernalization pathway (Kim and Sung, 2014) (Figure 18B).

4.3.4. Hormonal imbalance plays a role in sugarcane smut disease

Changes in expression profiles of genes related to regulation, synthesis, and transport of hormones identified in the enrichment analysis were investigated. We detected that JA (jasmonic acid)-mediated as well as SA (salicylic acid)-signaling are potentially restrained in the experiments 5-DAI (Figure 17). Indeed Que et al. (2014) using a smut resistant sugarcane genotype identified overexpression of JA-associated genes.

In contrast to SA and JA, auxin-activated signaling pathways are up-regulated. In addition to acting as a negative regulator of the plant immune system (Robert-Seilaniantz *et al.*, 2007; Wang *et al.*, 2007), we suspect that the auxin-related DEGs are associated with the meristem transcriptional reprogramming during whip emission, since an increased auxin transporter gene expression was confirmed by RT-qPCR in both 5 and 200-DAI samples (Figure 19). The balance of auxin-cytokinin is reported as essential for typical meristem function (Su *et al.*, 2011). Additionally, auxin is the hormone with the highest number of responsive genes up-regulated after whip emission (Figure 20), including those involved in auxin influx/efflux, auxin-amino acid hydrolase, and auxin-responsive proteins such as Aux/IAA, SAUR, and auxin-induced β-glucosidase. In the *S. reilianum*-maize pathosystem, the floral reversion process is partially attributed to an increase in auxin concentration, contributing to the loss of apical dominance and a greater number of ears per branch (Ghareeb *et al.*, 2011). An auxin-dependent signaling is likely necessary for whip emission, which also involves the loss of apical dominance and growth of secondary buds (Sundar *et al.*, 2012). Increase in auxin-like substances were found in Hector et al. (1992) work using smutted sugarcane extracts. The authors suggested that the balance between auxin and cytokinin is disrupted in infected sugarcane plants.

Although cytokinin is a hormone often related to tillering (Liu *et al.*, 2011), a symptom often related to smut, genes related to its synthesis were not detected among those that were differentially expressed (Figure 20). However, several type-A response regulators (RRs) were up-regulated 200 DAI. Type-A RRs negatively regulate cytokinin signaling by repressing type-B RRs, and they are transcriptionally up-regulated in response to cytokinin (To *et al.*, 2007). Cytokinins are central regulators in the maize smut caused by *U. maydis*. The fungus is able to synthesize

cytokinins, an important virulence factor that is associated with uncoordinated cell division and tumor formation (Brefort *et al.*, 2014; Bruce *et al.*, 2011). The *S. scitamineum* genome SSC39B (Taniguti *et al.*, 2015) does not have cytokinin biosynthetic gene homologs. However, previously data stated that *S. scitamineum* secretes some cyokinin-like substances potentially able to activate cytokinin-responsive genes (Peros and Chagvardieff, 1983).

Regarding ethylene (ET), DEGs related to biosynthesis, perception, and signal transduction were detected after whip emission (Figure 20). For instance, SAM (S-adenosyl-L-methionine) was identified as up-regulated and confirmed by RT-qPCR analysis (Figure 19). ET is often related to lignification of plant tissues by increasing the expression of genes involved in the phenylpropanoid pathway (Ecker and Davis, 1987; Guo and Ecker, 2004).

The genes related to brassinosteroids were also up-regulated. They included several BAK1 LRRs (Li *et al.*, 2002) and the transcription repressor BZR1, which binds directly to the promoters)of feedback-regulated brassinosteroid biosynthetic genes (He *et al.*, 2005). The identification of GA2ox and DELLA proteins among the DEGs suggests a blockage of GA signaling, because the products of these genes act by reducing the availability of active GAs and repressing GA-responsive genes, respectively (Lo *et al.*, 2008; Thomas *et al.*, 1999) (Figure 20). For instance, rice plants overexpressing C20-GA2ox exhibit early tillering, adventitious root growth, and changes in plant architecture that generate semi-dwarfs (Lo *et al.*, 2008). Diseased sugarcane plants have reduced node distances and poor cane formation (Sundar *et al.*, 2012, our own observations), similar to the symptoms described in the rice with low GA levels.

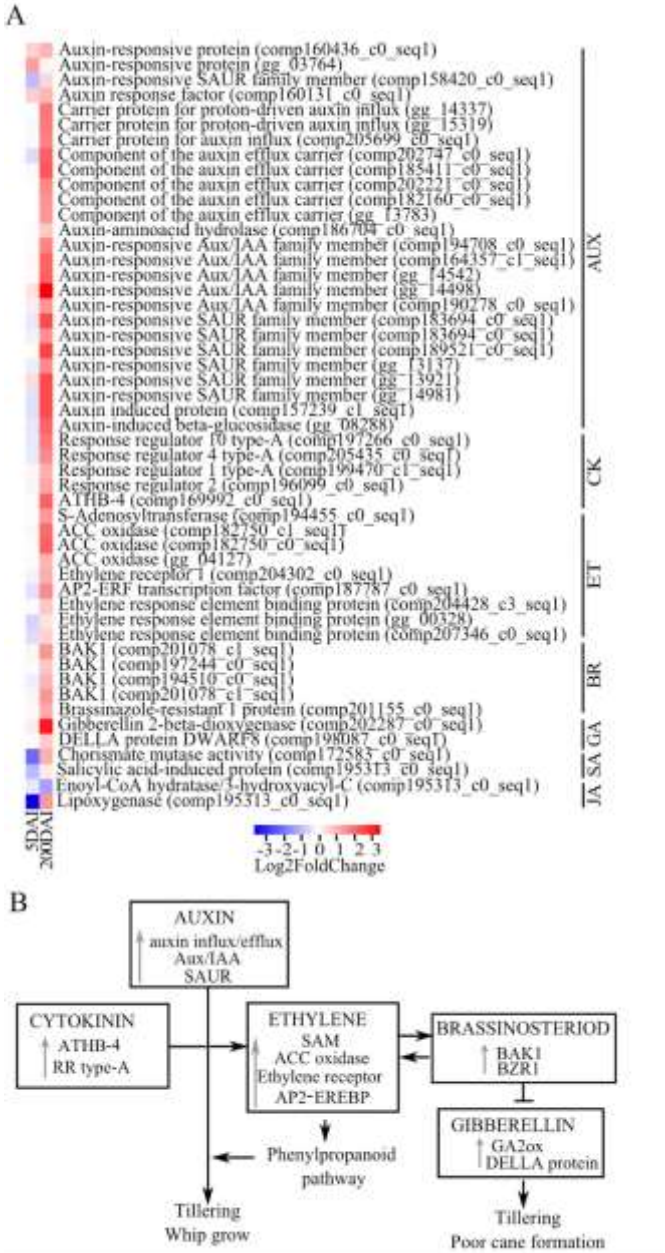


Figure 20. DEGs related to hormone biosynthesis and signalization. (A) Heatmap constructed in R software package using the Log2 Fold Change values (Inoculated/Control). Left column represents regulation of each gene at 5 DAI and right column at 200 DAI. (B) Model concerning the contribution of the main plant hormones with up-regulated DEGs in sugarcane after whip emission.

4.3.5. Carbon distribution is affected by *S. scitamineum* colonization

In biotrophic interactions pathogen growth relies on host nutrients derived from an active metabolism. Sucrose and its derivatives are central molecules involved in carbohydrate translocation, metabolism, and sensing in higher plants (Ruan, 2014). Invertases were among the DEGs up-regulated after whip emission and included a neutral alkaline invertase (*comp189016_c0_seq1*) and a soluble acid invertase (*comp201528_c0_seq1*), both confirmed by RT-

qPCR analysis. Invertases catalyze the irreversible hydrolysis of sucrose (EC 3.2.1.26), in some cases leading to a shift of the apoplastic sucrose/hexose ratio in favor of the hexoses (Tauzin and Giardina, 2014), which regulate many aspects of plant metabolism such as carbohydrate portioning, developmental processes, and hormonal responses to biotic stress (Roitsch and González, 2004; Tauzin and Giardina, 2014). Plant invertases are classified in three groups: alkaline/neutral invertases localized in the cytosol, mitochondria or plastids; and two types of acid invertases, one insoluble and bound to the cell wall (cell wall invertase CWI), and the other soluble in the vacuole space (vacuolar invertase VI) (Roitsch and González, 2004; Tauzin and Giardina, 2014). The acid invertases CWI play a role in sucrose partitioning, plant development and cell differentiation, whereas the VIs determine the sucrose level stored in the vacuole and its remobilization for metabolic processes. The up-regulation of vacuolar and neutral invertases represents a shift in the plant's metabolism that targets carbon to pathways unrelated to sucrose storage, which can aggravate sugarcane-smut symptoms. Additionally, it has relevant implications for the hexose-based sugar signaling system involved in plant immunity (Morkunas and Ratajczak, 2014). Increased hexose levels can also be related to the nutrients supplied to the pathogen during teliospore differentiation and whip formation (Figure 21). The importance of the sugar content in the signaling for axillary bud growth was recently demonstrated and indicates that in addition to auxin, an increased sugar supply is necessary and sufficient for suppressed buds to be released from apical dominance (Mason *et al.*, 2014). Redistribution of the host carbon in response to *S. scitamineum* sporogenesis is suggested by the transcriptional profiles of the genes related to glycolysis, the citric acid cycle, sucrose, starch, xylan, trehalose 6P, and cellulose biosynthesis (Figure 21), also confirmed by RT-qPCR analysis (Figure 19). It has been suggested that rather than playing a metabolic role, the low concentration of trehalose-6P (T6P) in infected plants functions as a regulatory component. Trehalose-6P synthase can sense sucrose availability to generate T6P as a signal to promote growth (Gómez *et al.*, 2006; Kolbe *et al.*, 2005; Ruan, 2014). There is also evidence in *A. thaliana* that T6P acts as an endogenous signal to control the transition from vegetative growth to flowering by increasing trehalose-6P synthase transcript levels (Ruan, 2014; Wahl *et al.*, 2013). These findings should encourage new experiments to better understand the sugarcane metabolic response to smut as the disease progresses, the turning point at which the plant changes its metabolism to allow fungal sporogenesis, and the significance of this shift to teliospore/whip differentiation.

Several transcripts related to lignin biosynthesis were detected as up-regulated after whip emission such as shikimate hydroxycinnamoyl transferase (EC 2.3.1.133), cinnamoyl-CoA reductase (EC 1.2.1.44), and peroxidase (EC 1.11.1.7). In addition, up-regulated laccases were

detected also as DEGs (*gg_01080, gg_14238*, *gg_10439, gg_15488*). Plant laccases (EC 1.10.3.2) are glycoproteins involved in lignin biosynthesis through the oxidation of lignin precursors (Dean *et al.*, 1998). An increase in the lignification of smut-resistant plants has been detected by measuring cinnamyl alcohol dehydrogenase levels and by the overexpression of genes in RNAseq experiments of resistant varieties in the early moments of interaction (Que, Su, *et al.*, 2014; Santiago *et al.*, 2012). The increase in lignin content after whip emission is probably not related to a protective host response, but instead, is likely a stage in the formation of the whip, which is composed of lignified plant tissue (Legaz *et al.*, 2011). Recently, a proteomic approach developed after whip emission revealed 53 proteins related to lignin accumulation and oxidative stress at this stage of disease symptoms (Barnabas *et al.*, 2016). Responses regarding ROS (*Reactive Oxygen Stress*) in 200-DAI samples were also detected at RNA level. Nineteen DEs were identified as related to ROS-scavenging enzymes, including 16 up-regulated DEGs (one catalase, eight peroxidases, two thioredoxins, and five glutathione S-transferases) and three down-regulated DEGs (two peroxidases and one thioredoxin). These results suggest that the ROS level is high in *S. scitamineum*-colonized cells during sporogenesis.

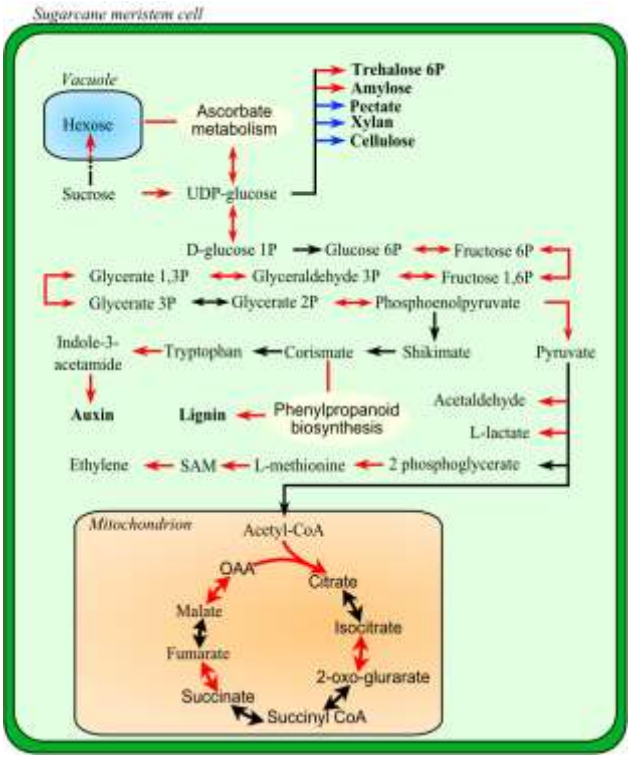


Figure 21. Smutted sugarcane metabolism in late moments of interaction. Schematic representation of smutted sugarcane metabolism in the whip base (200 DAI). Red and blue arrows represent up- and down-regulated, respectively, and black arrows are unchanged expression.

4.3.6. Sequence features of resistance gene analogs (RGAs) differentially expressed

Resistance gene analogs (RGAs) were analyzed for both time points. Predicted domains and other sequence features that are potentially important for RGAs function were identified (Figure 22). Although this study was conducted with an intermediate resistant genotype, we detected promising candidates associated with this particular biotrophic interaction and their potential role in the disease progression mechanism proposed here. Several RGAs containing leucine-rich repeat (LRR) domains have already been identified in sugarcane (Rossi *et al.*, 2003). Three of them were also found here: the two BAM-related proteins *comp_188744* and *gg_06875* (RGA482) and a protein encoded by *comp_187876* (RGA367) (Figure 22). All the other proteins predicted in this work are new discoveries in sugarcane.

BAM orthologs are leucine-rich repeat receptor-like serine/threonine-RD kinases (LRR-RLKs) (Afzal *et al.*, 2008) and receptors for signals to switch meristem identity. BAM proteins have roles equivalent to that of CLV (**CL**A**V**ATA) proteins in *A. thaliana*, representing a functional redundancy within the program related to meristem functionality (DeYoung *et al.*, 2006). Other LRR-RLKs such as ERECTA were first described in relation to plant development, and only later were they positively associated with disease resistance (Afzal *et al.*, 2008).

Other kinases were identified sharing amino acid identities with RLKs of different families (Figure 22). For instance, *comp205382_c0_seq1* encodes a transmembrane protein (non-RD kinase) harboring an LRR domain (Afzal *et al.*, 2008). The translated amino acid sequence is most similar to the protein Xa21 in *O. sativa*, sharing all conserved residues (Song *et al.*, 1995). The *xa21* gene is known to confer resistance to *Xanthomonas oryzae* pv. *oryzae* race 6 in rice (Song *et al.*, 1995). LRR-RLK-harboring domains shared by plant S-locus glycoproteins and S-receptor kinases (SRK) (IPR003609) were also detected (*comp182409_c0_seq1*). SRKs were first described as being associated with *Brassica* self-incompatibility, but were later related to the perception of pathogen infection, probably by binding to a glycoprotein inducer such as cellulose or chitin (Pastuglia *et al.*, 1997). A transcript (EST) similar to *comp182409_c0_seq1* was detected by BLAST (GenBank) in *Oryza longistaminata*. *Comp182409_c0_seq1* shares an 87% identity to the 3' end of the *xa21* nucleotide sequence (FF359116), which is the portion coding for the kinase domain. No sugarcane ESTs similar to *comp182409_c0_seq1* were found in the NCBI expressed sequence tags database.

S. scitamineum also induced the presence of cytoplasmic LRR proteins known as plant intracellular *Ras* group-related LRR proteins (PIRLs) (Forsthoefel *et al.*, 2005). The protein encoded by the *comp196006_c0_seq1* DEG at 200 DAI is most similar to other PIRL4s and

contains the conserved GxxxVxxYxxxxW ('GVYW') motif immediately following the LRR domain.

The expression of an F-box/LRR related gene was also detected in 200-DAI samples. F-box proteins are part of the SCF (SKP1/Cullin/F-box) ubiquitin ligase complex involved in protein degradation (proteasome) (Jain *et al.*, 2007). In *O. sativa,* the F-box/LRR-repeat MAX2-homolog controls tillering by suppressing axillary bud activity, potentially by degrading specific proteins that activate axillary growth (Ishikawa *et al.*, 2005). The *comp204415_c0_seq1* DEG encoded a protein that has an 83% amino acid identity with the rice MAX2-homolog (Q5VMP0). Tillering is one of the earlier disease symptoms related to smut (Sundar et al., 2012).

Cytoplasmic proteins attached to the nucleotide binding-ARC (NB-ARC) domains and containing an ATPase and a nucleotide-binding site (McHale *et al.*, 2006; van Ooijen *et al.*, 2008) compose the last class of LRR proteins identified among the translated transcripts that were differentially accumulated. The translated protein sequence of *comp202536_c0_seq1* contains the conserved GLPLA and MHD motifs essential to the function of other resistance proteins (van Ooijen *et al.*, 2008), and although lacking an obvious CC-domain at the N-terminal, it probably belongs to one of the CC-domain-containing subfamilies. The sequence is most similar to an *O. sativa* gene assigned to chromosome 4 (CAE03396). All these RGAs-like encoding proteins are potential targets for functional characterization as receptors of signals due to the presence of *S. scitamineum*.

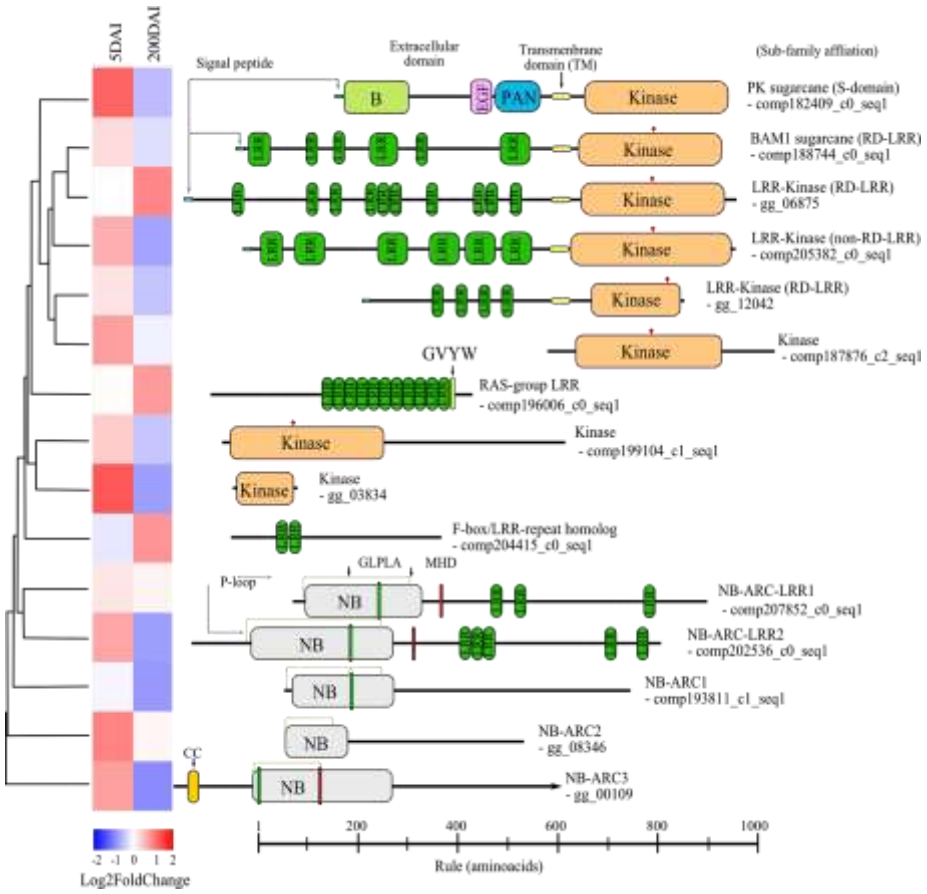


Figure 22. Resistance gene analogs (RGAs) detected as DEGs in smut infected sugarcane in early and late interaction. Guide tree was obtained based on translated amino acid sequence similarity using CLUSTALW2. Heatmap represent their respective expression Log2 fold change values. Protein structural features: Signal peptide sequences predicted by SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/); Transmembrane domains predicted by TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). Other domains, motifs and active sites were predicted by InterProScan. CC, coiled coil; kinase (IPR011009); NB, nucleotide-binding (IPR002182); B, Bulb-type lectin domain (IPR001480); EGF, epidermal growth factor domain (IPR000742); PAN domain (IPR03609), LRR, leucine-rich-repeat (IPR001611; IPR003591); red pins represent kinase active sites; and gray arrows are P-loops of N-terminal NB-ARC proteins.

4.4. Conclusion

This work reveals transcriptional changes associated with the most characteristic symptom of sugarcane smut disease. We speculate that the whip emission is a consequence of premature transcriptional changes in meristem function (5-DAI) that results in restraining of floral development via vernalization pathway by increasing VIN3, COL6, FT, and AP1 gene expression and other flowering-related transcriptional factors (200-DAI). The fungal sporogenesis and whip emission are most related to auxin mobilization followed by a strong response of ROS scavenging enzymes. In addition, the role of other plant hormones is also suggested. Because the processes associated with fungal development and whip emission require energy, carbon partitioning of sugarcane is the most affected. Gene expression profile indicates that smutted sugarcane metabolism shifts towards energy production, increasing the expression of genes

involved in glycolysis and TCA. Synthesis of signalizing molecules such as trehalose 6P is also among the results described. Increased expression of genes involved in lignin biosynthesis and sucrose breakdown are potential markers of the whip development. We also described RGAs expression patterns involved in this particular interaction leading to an effective fungal colonization and disease establishment. *S. scitamineum* is known to colonize not only susceptible plants but also smut resistant genotypes that in response to unknown signals allow unexpectedly fungal sporogenesis and the whip emission. This detailed work is an attempt to expose molecular mechanisms and candidate genes that can possibly reveals ways to control sugarcane smut disease.

4.5. Database accession number

The sequencing data has been deposited at DDBJ/EMBL/GenBank under the BioProject ID PRJNA291816.

REFERENCES

Afzal, A.J., Wood, A.J. and Lightfoot, D.A. (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Mol. Plant-Microbe Interact.* **21**, 507–517.

Bakkeren, G., Kämper, J. and Schirawski, J. (2008) Sex in smut fungi: structure, function and evolution of mating-type complexes. *Fungal Genet. Biol.* **45**, 15–21.

Barnabas, L., Ashwin, N.M.R., Kaverinathan, K., et al. (2016) Proteomic analysis of a compatible interaction between sugarcane and *Sporisorium scitamineum*. *Proteomics* **16**, 1111–22.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. and Wheeler, D.L. (2000) GenBank. *Nucleic Acids Res.* **28**, 15–18.

Borrás-Hidalgo, O., Thomma, B.P.H.J., Carmona, E., Borroto, C.J., Pujol, M., Arencibia, A. and Lopez, J. (2005) Identification of sugarcane genes induced in disease-resistant somaclones upon inoculation with *Ustilago scitaminea* or *Bipolaris sacchari*. *Plant Physiol. Biochem.* **43**, 1115–1121.

Brefort, T., Tanaka, S., Neidig, N., Doehlemann, G., Vincon, V. and Kahmann, R. (2014) Characterization of the largest effector gene cluster of *Ustilago maydis* Xu, J.-R., ed. *PLoS Pathog.* **10**, e1003866.

Bruce, S., Saville, B. and Neil Emery, R.J. (2011) *Ustilago maydis* produces cytokinins and abscisic acid for potential regulation of tumor formation in maize. *J. Plant Growth Regul.* **30**, 51–63.

Bueno, C.R.N.C. (2010) *Infection by Sporisorium scitamineum on sugarcane: influence of environmental variables and development of a method for early diagnosis*. Universidade de São Paulo.

Cardoso-Silva, C.B., Costa, E.A., Mancini, M.C., et al. (2014) De novo assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS One* **9**, e88462.

Castro, E. de, Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365.

Chao CP, Hoy JW, Saxton AM, M.F. (1990) Heritability of resistance and repeatability of clone reactions to sugarcane smut in Louisiana. *Phytopathology* **80**, 622–626.

Conesa, A. and Götz, S. (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**, 619832.

Consortium, T.U. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212.

D'Hont, A., Grivet, L., Feldmann, P., Glaszmann, J.C., Rao, S. and Berding, N. (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol. Gen. Genet.* **250**, 405–413.

Dean, J.F.D., LaFayette, P.R., Rugh, C., Tristram, A.H., Hoopes, J.T., Eriksson, K.-E.L. and Merkle, S.A. (1998) Laccases associated with lignifying vascular tissues. In Lignin and Lignan Biosynthesis., ACS Symposium Series. pp. 8–96. American Chemical Society.

DeYoung, B.J., Bickle, K.L., Schrage, K.J., Muskett, P., Patel, K. and Clark, S.E. (2006) The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. *Plant J.* **45**, 1–16.

Eckardt, N.A. (2010) Dissecting cis-Regulation of FLOWERING LOCUS T. *Plant Cell* **22**, 1422.

Ecker, J.R. and Davis, R.W. (1987) Plant defense genes are regulated by ethylene. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 5202–5206.

FAO (2013) Food and Agriculture Organization of United Nations - FAOSTAT. *Production*.

Forsthoefel, N.R., Cutler, K., Port, M.D., Yamamoto, T. and Vernon, D.M. (2005) PIRLs: a novel class of plant intracellular leucine-rich repeat proteins. *Plant Cell Physiol.* **46**, 913–22.

Gao, L., Kelliher, T., Nguyen, L. and Walbot, V. (2013) *Ustilago maydis* reprograms cell proliferation in maize anthers. *Plant J.* **75**, 903–14.

Garcia, A.A.F., Mollinari, M., Marconi, T.G., et al. (2013) SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci. Rep.* **3**, 3399.

Ghareeb, H., Becker, A., Iven, T., Feussner, I. and Schirawski, J. (2011) *Sporisorium reilianum* infection changes inflorescence and branching architectures of maize. *Plant Physiol.* **156**, 2037–2052.

Gómez, L.D., Baud, S., Gilday, A., Li, Y. and Graham, I.A. (2006) Delayed embryo development in the ARABIDOPSIS TREHALOSE-6-PHOSPHATE SYNTHASE 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. *Plant J.* **46**, 69–84.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech* **29**, 644–652.

Guo, H. and Ecker, J.R. (2004) The ethylene signaling pathway: new insights. *Curr. Opin. Plant Biol.* **7**, 40–49.

He, J.-X., Gendron, J.M., Sun, Y., Gampala, S.S.L., Gendron, N., Sun, C.Q. and Wang, Z.-Y. (2005) BZR1 is a transcriptional repressor with dual roles in brassinosteroid homeostasis and growth responses. *Science (80-.).* **307**, 1634–1638.

Hector E, Rodriguez R, de Prada F, Delmonte A, G.R. (1992) Experimental evidente dor the presence of differente smut resistance mechanisms in sugarcane. In XXI Congress of the International Society of Sugarcane Technologists. (Napompeth, B., ed), pp. 565–573. Bankok: International Society of Sugarcane Technologists.

Honma, T. and Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**, 525–529.

Ishikawa, S., Maekawa, M., Arite, T., Onishi, K., Takamure, I. and Kyozuka, J. (2005) Suppression of tiller bud activity in tillering dwarf mutants of rice. *Plant Cell Physiol.* **46**, 79–86.

Iskandar, H., Simpson, R., Casu, R., Bonnett, G., Maclean, D. and Manners, J. (2004) Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. *Plant Mol. Biol. Report.* **22**, 325–337.

Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., Kapoor, S., Tyagi, A.K. and Khurana, J.P. (2007) F-Box Proteins in Rice. Genome-Wide Analysis, Classification, Temporal and Spatial Gene Expression during Panicle and Seed Development, and Regulation by Light and Abiotic Stress. *Plant Physiol.* **143**, 1467–1483.

James, G.L. (1973) Smut spore germination in sugarcane internode surfaces. *Proc. South African Sugar Technol. Assoc.*, 179–180.

Jones, P., Binns, D., Chang, H.-Y., et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.

Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30.

Kaufmann, K., Wellmer, F., Muiño, J.M., et al. (2010) Orchestration of floral initiation by APETALA1. *Science (80-.).* **328**, 85–89.

Kawahara, Y., Oono, Y., Kanamori, H., et al. (2012) Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. Lee, Y.-H., ed. *PLoS One* **7**, e49423.

Kim, D.-H. and Sung, S. (2014) Genetic and epigenetic mechanisms underlying vernalization. *Arabidopsis Book* **12**, e0171.

Kolbe, A., Tiessen, A., Schluepmann, H., Paul, M., Ulrich, S. and Geigenberger, P. (2005) Trehalose 6-phosphate regulates starch synthesis via posttranslational redox activation of ADP-glucose pyrophosphorylase. *Proc. Natl. Acad. Sci. United States Am.* **102**, 11118–11123.

Krogh, A., Larsson, B., Heijne, G. von and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–80.

Lagercrantz, U. and Axelsson, T. (2000) Rapid Evolution of the Family of CONSTANS LIKE Genes in Plants. *Mol. Biol. Evol.* **17**, 1499–1507.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.

LaO, M., Arencibia, A., Carmona, E., Acevedo, R., Rodríguez, E., León, O. and Santana, I. (2008) Differential expression analysis by cDNA-AFLP of *Saccharum* spp. after inoculation with the host pathogen *Sporisorium scitamineum*. *Plant Cell Rep.* **27**, 1103–1111.

Lee, Y.K., Kim, G.-T., Kim, I.-J., Park, J., Kwak, S.-S., Choi, G. and Chung, W.-I. (2006) LONGIFOLIA1 and LONGIFOLIA2, two homologous genes, regulate longitudinal cell elongation in *Arabidopsis*. *Development* **133**, 4305–4314.

Legaz, M.E., Armas, R. d., Pinon, D. and Vicente, C. (1998) Relationships between phenolics-conjugated polyamines and sensitivity of sugarcane to smut (*Ustilago scitaminea*). *J. Exp. Bot.* **49**, 1723–1728.

Legaz ME, Santiago R, de Armas R, Alarcón B, Díaz EM, Sánchez-Elordi E, Sacristán M, Fontaniella B, Millanes AB, B.M., Legaz, M.E., Santiago, R., Armas, R. De, Millanes, a M., Blanch, M. and Vicente, C. (2011) Molecular defence responses of sugarcane (*Saccharum officinarum* L.) to smut (*Sporisorium scitamineum* (Syd.). In Science against microbial pathogens: communicating current research and technological advances. (A, M.-V., ed), pp. 1244–1250. Piepenbr & Oberw.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li, J., Wen, J., Lease, K.A., Doke, J.T., Tax, F.E. and Walker, J.C. (2002) BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **110**, 213–222.

Liu, Y., Ding, Y., Gu, D., Li, G., Wang, Q. and Wang, S. (2011) The positional effects of auxin and cytokinin on the regulation of rice tiller bud growth. *Crop Sci.* **51**, 2749–2758.

Lloyd, H.L. and Naidoo, G. (1983) Chemical-assay potentially suitable for determination of smut resistance of sugarcane cultivars. *Plant Dis.* **67**, 1103–1105.

Lloyd, H.L. and Pillay, M. (1980) The development of an improved method for evaluating sugarcane for resistance to smut. *Proc. South African Sugar Technol. Assoc.* **54**, 168–172.

Lo, S.-F., Yang, S.-Y., Chen, K.-T., Hsing, Y.-I., Zeevaart, J.A.D., Chen, L.-J. and Yu, S.-M. (2008) A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *Plant Cell* **20**, 2603–2618.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Lowe, R.G.T., Cassin, A., Grandaubert, J., Clark, B.L., Wouw, A.P. Van de, Rouxel, T. and Howlett, B.J. (2014) Genomes and transcriptomes of partners in plant-fungal- interactions between canola (Brassica napus) and two *Leptosphaeria species*. *PLoS One* **9**, e103098.

Mason, M.G., Ross, J.J., Babst, B.A., Wienclaw, B.N. and Beveridge, C.A. (2014) Sugar demand, not auxin, is the initial regulator of apical dominance. *Proc. Natl. Acad. Sci.* **111**, 6092–6097.

McHale, L., Tan, X., Koehl, P. and Michelmore, R.W. (2006) Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* **7**, 212.

Morkunas, I. and Ratajczak, L. (2014) The role of sugar signaling in plant defense responses against fungal pathogens. *Acta Physiol. Plant.* **36**, 1607–1619.

Murai, K. (2013) Homeotic genes and the ABCDE model for floral organ formation in wheat. *Plants* **2**, 379–395.

Ooijen, G. van, Mayr, G., Albrecht, M., Cornelissen, B.J.C. and Takken, F.L.W. (2008) Transcomplementation, but not physical association of the CC-NB-ARC and LRR domains of tomato R protein Mi-1.2 is altered by mutations in the ARC2 subdomain. *Mol. Plant* **1**, 401–10.

Palhares, A.C., Rodrigues-Morais, T.B., Sluys, M.-A. Van, et al. (2012) A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. *BMC Genet.* **13**, 51.

Papini-Terzi, F.S. (2005) Transcription profiling of signal transduction-related genes in sugarcane tissues. *DNA Res.* **12**, 27–38.

Pastuglia, M., Ruffio-Châble, V., Delorme, V., Gaude, T., Dumas, C. and Cock, J.M. (1997) A functional S locus anther gene is not required for the self-incompatibility response in *Brassica oleracea*. *Plant Cell* **9**, 2065–2076.

Peros, J.P. and Chagvardieff, P. (1983) Mise en evidence in vitro d'effets toxiques entre *Ustilago scitaminea* Syd. et des cats de canne a sucre; influence de deux facteurs de croissance sur les phenomenes observes. *Agronomie* **3**, 629–634.

Petersen, T.N., Brunak, S., Heijne, G. von and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth* **8**, 785–786.

Pfaffl, M.W., Horgan, G.W. and Dempfle, L. (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36–e36.

Piñon, D., Armas, R. de, Vicente, C. and Legaz, M.E. (1999) Role of polyamines in the infection of sugarcane buds by *Ustilago scitaminea* spores. *Plant Physiol. Biochem.* **37**, 57–64.

Que, Y., Su, Y., Guo, J., Wu, Q. and Xu, L. (2014) A global view of transcriptome dynamics during *Sporisorium scitamineum* challenge in sugarcane by RNA-seq Fu, B., ed. *PLoS One* **9**, e106476.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L., Moorman, A.F.. M., Lekanne Deprez, R.H. and Moorman, A.F.. M. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66.

Robert-Seilaniantz, A., Navarro, L., Bari, R. and Jones, J.D.G. (2007) Pathological hormone imbalances. *Curr. Opin. Plant Biol.* **10**, 372–9.

Roitsch, T. and González, M.-C. (2004) Function and regulation of plant invertases: sweet sensations. *Trends Plant Sci.* **9**, 606–613.

Rossi, M., Araujo, P.G., Paulet, F., Garsmeur, O., Dias, V.M., Chen, H., Sluys, M.-A. Van and D'Hont, A. (2003) Genomic distribution and characterization of EST-derived resistance gene analogs (RGAs) in sugarcane. *Mol. Genet. Genomics* **269**, 406–419.

Ruan, Y.-L. (2014) Sucrose metabolism: gateway to diverse carbon use and sugar signaling. *Annu. Rev. Plant Biol.* **65**, 33–67.

Santiago, R., Alarcón, B., Armas, R. de, Vicente, C. and Legaz, M.E. (2012) Changes in cinnamyl alcohol dehydrogenase activities from sugarcane cultivars inoculated with *Sporisorium scitamineum* sporidia. *Physiol. Plant.* **145**, 245–259.

Setta, N. de, Monteiro-Vitorello, C.B., Metcalfe, C.J., et al. (2014) Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* **15**, 540.

Simon, R., Igeno, M.I. and Coupland, G. (1996) Activation of floral meristem identity genes in Arabidopsis. *Nature* **384**, 59–62.

Song, W.-Y., Wang, G.-L., Chen, L.-L., et al. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science (80-.).* **270**, 1804–1806.

Su, Y.-H., Liu, Y.-B. and Zhang, X.-S. (2011) Auxin–cytokinin interaction regulates meristem development. *Mol. Plant* **4**, 616–625.

Sundar, AR; Barnabas, EL; Malathi, P; Viswanathan, R., Sundar, A.R., Barnabas, E.L., Malathi, P. and Viswanathan, R. (2012) A Mini-Review on smut disease of sugarcane caused by *Sporisorium scitamineum*. In Botany. (Mworia, J.K., ed), pp. 107–128. InTech.

Taniguti, L.M., Schaker, P.D.C., Benevenuto, J., et al. (2015) Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with Sugarcane. *PLoS One* **10**, e0129318.

Tauzin, A.S. and Giardina, T. (2014) Sucrose and invertases, a part of the plant defense response to the biotic stresses. *Front. Plant Sci.* **5**, 293.

Thokoane, L.N. and Rutherford, R.S. (2001) cDNA-AFLP differential display of sugarcane (*Saccharum* spp. hybrids) genes induced by challenge with the fungal pathogen *Ustilago scitaminea* (sugarcane smut). *Proc. South African Sugar Technol. Assoc.* **75**, 104–107.

Thomas, S.G., Phillips, A.L. and Hedden, P. (1999) Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proc. Natl. Acad. Sci.* **96**, 4698–4703.

To, J.P.C., Deruère, J., Maxwell, B.B., Morris, V.F., Hutchison, C.E., Ferreira, F.J., Schaller, G.E. and Kieber, J.J. (2007) Cytokinin regulates type-A Arabidopsis response regulator activity and protein stability via two-component phosphorelay. *Plant Cell* **19**, 3901–3914.

Tremblay, A., Hosseini, P., Alkharouf, N.W., Li, S. and Matthews, B.F. (2011) Gene expression in leaves of susceptible *Glycine max* during infection with *Phakopsora pachyrhizi* using next generation sequencing. *Sequencing* **2011**, 1–14.

Waclawovsky, A.J., Sato, P.M., Lembke, C.G., Moore, P.H. and Souza, G.M. (2010) Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. *Plant Biotechnol. J.* **8**, 263–76.

Wahl, V., Ponnu, J., Schlereth, A., et al. (2013) Regulation of flowering by trehalose-6-phosphate signaling in *Arabidopsis thaliana*. *Science (80-.).* **339**, 704–707.

Wang, D., Pajerowska-Mukhtar, K., Culler, A.H. and Dong, X. (2007) Salicylic acid inhibits pathogen growth in plants through repression of the auxin signaling pathway. *Curr. Biol.* **17**, 1784–90.

Westermann, A.J., Gorski, S.A. and Vogel, J. (2012) Dual RNA-seq of pathogen and host. *Nat Rev Micro* **10**, 618–630.

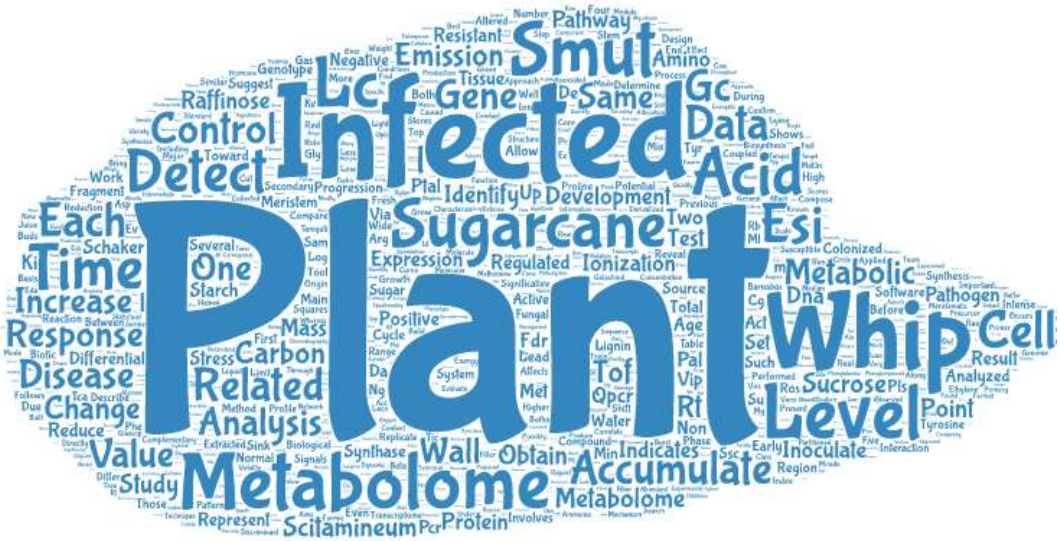
Wu, Q., Xu, L., Guo, J., Su, Y. and Que, Y. (2013) Transcriptome profile analysis of sugarcane responses to *Sporisorium scitaminea* infection using solexa sequencing technology. *Biomed Res. Int.* **2013**, 1–9.

Yazawa, T., Kawahigashi, H., Matsumoto, T. and Mizuno, H. (2013) Simultaneous transcriptome analysis of *Sorghum* and *Bipolaris sorghicola* by using RNA-seq in combination with de novo transcriptome assembly. *PLoS One* **8**, e62460.

You-Xiong, Q., Jian-Wei, L., Xian-Xian, S., Li-Ping, X. and Ru-Kai, C. (2011) Differential gene expression in sugarcane in response to challenge by fungal pathogen *Ustilago scitaminea* revealed by cDNA-AFLP. *J. Biomed. Biotechnol.* **2011**, 1–10.

Zhu, Q.-H., Stephen, S., Kazan, K., Jin, G., Fan, L., Taylor, J., Dennis, E.S., Helliwell, C.A. and Wang, M.-B. (2013) Characterization of the defense transcriptome responsive to Fusarium oxysporum-infection in Arabidopsis using RNA-seq. *Gene* **512**, 259–66.

5. ASSESSING THE METABOLIC CLUTTER IN A SUGARCANE-*SPORISORIUM SCITAMINEUM* COMPATIBLE INTERACTION FROM EARLY INFECTION TO WHIP DEVELOPMENT



Abstract

In response to pathogen attack, plants modulate gene expression culminating in metabolic reprogramming, which may be directly related to resistance or susceptibility phenotypes. Sugarcane smut disease, caused by the biotrophic fungus *Sporisorium scitamineum*, is characterized by the development of a structure like a whip originated from the plant meristem where fungal sporulation occurs. The disease causes negative effects on sucrose accumulation, fiber content and juice quality. These global responses are the result of extensive metabolic reprogramming to sustain pathogen colonization and whip growth. The objective of this study was to obtain metabolic profiles of the smut-sugarcane pathosystem since *S. scitamineum* infection, throughout disease progression until the whip emission. The complementary techniques CG-MS and LC-MS were used combined with previous obtained transcriptomic data. The main results showed that shortly after infection occurred a significant increase of raffinose levels. Disease progression was characterized by a shift in the primary metabolism between 65 and 100 DAI, especially of those metabolites related to cell wall biosynthesis. These changes leaned toward reduced components of the cell wall at 100 and 120 DAI, suggesting loosening of the cell wall to allow whip growth. Energetic pathways and possibly starch accumulation were affected throughout the interaction. It is suggested that increased levels of tyrosine in infected plants may be related of differential PTAL gene expression possibly leading to the synthesis of lignin, whereas altered levels of methionine may be related to ethylene hormonal imbalance. Unique secondary metabolites, including a fungal toxin, were detected since 65 DAI, and shall be exploited as biomarkers of sugarcane smut disease. This is the first report of the use of the metabolomics technique to understand a sugarcane pathosystem, and provides valuable information to the study of smut resistance in sugarcane.

5.1. Introduction

Sugarcane (*Saccharum* spp.) has long been recognized as one of the world's most efficient crop in converting solar into harvestable chemical energy, and for this reason it is the main feedstock for sugar and ethanol production in tropical and subtropical countries (Botha and Moore, 2014). Sugarcane stores exceptionally high concentrations of sucrose, which can achieve 25% of fresh weight under favorable conditions (Chandra, 2011). The carbon partitioning is directly related to the well-established concept of source (photosynthetic) and sink (non-photosynthetic) tissues systems (McCormick *et al.*, 2009). Sucrose synthesized in source tissues is transported via phloem, distributed via apoplast (Robinson-Beers and Evert, 1991) and symplast (Rae *et al.*, 2005). Typically, in sugarcane mature tissues, carbon skeletons are converted to sucrose and stored in cellular vacuoles, whereas in younger tissues sucrose is used to build proteins and to cell wall fiber synthesis (Bindon and Botha, 2002; Rae *et al.*, 2005).

In the sucrose cycle, carbon is partitioned into several compounds including organic acids, amino acids, proteins, cell wall components and secondary metabolites (Botha and Whittaker, 1997; Wang *et al.*, 2013). In response to pathogen attack, carbon partitioning can be affected by the activation of a wide range of defense mechanisms, which involves the redistribution of energy to the synthesis of secondary metabolites, cell wall reinforcement, production of reactive oxygen species (ROS) and changes in hormonal status (Bolton, 2009). Sugarcane is constantly challenged by biotic stress, which can compromise crop productivity. The sugarcane smut is one of the most important diseases, leading to economic losses due to reduction in sugar content and juice quality (Sundar *et al.*, 2012). Sugarcane smut is caused by the biotrophic basidiomycete *Sporisorium scitamineum*. The characteristic symptom of the disease is the development of whip-like structure composed of sugarcane tissues surrounded by the fungal sporogenesis (Sundar *et al.*, 2012). Smut whip acts as a sink tissue depending on the plant carbon supply to grow (Doidy *et al.*, 2012). Many studies have been carried out in order to identify the molecular basis of this disease searching for changes in gene expression, protein accumulation and specific cell wall components that can be used as determinants of resistance (Barnabas *et al.*, 2016; Borrás-Hidalgo *et al.*, 2005; Esh, 2014; Fontaniella *et al.*, 2002; Heinze *et al.*, 2001; Huang *et al.*, 2015; LaO *et al.*, 2008; Millanes *et al.*, 2005; Piñon *et al.*, 1999; Que *et al.*, 2011; Santiago *et al.*, 2009; Santiago *et al.*, 2010; Santiago *et al.*, 2012; Schaker *et al.*, 2016; Su *et al.*, 2013; Taniguti *et al.*, 2015; You-Xiong *et al.*, 2011).

Changes in gene expression and/or protein accumulation are not always directly related to the observed biological function and phenotype, which are the result of a multiple regulatory

interactions (Fiehn *et al.*, 2000). Metabolomics has emerged as a complementary tool to functional genomics with the potential to accelerate the understanding of complex molecular interactions in biological systems (Hall *et al.*, 2002; Jorge *et al.*, 2016). This technique is one of the highest levels of post-genomic analysis, aiming to quantify compounds of intermediary metabolic pathways (Allwood *et al.*, 2008; Büscher *et al.*, 2009). Plant metabolome analysis is a great challenge because of the dynamic range of concentrations and the number of possible molecules, which can reach more than 200,000 variants (Fiehn *et al.*, 2000). Elucidation of this diversity is being achieved with the development and upgrading of analytical methods. Gas chromatography coupled to mass spectrometry (GC–MS) is the most widely accepted analytical method used in plant metabolomics due to its high reproducibility. One major limitation of GC–MS is the restriction to analysis of volatile and thermally stable metabolites or metabolites that can be chemically modified to produce volatile derivatives through derivatization. Liquid chromatography coupled to mass spectrometry (LC–MS) is the most important complementary technology to GC–MS, where thermolabile and high-molecular weight compounds without any derivatization can be analyzed (Jorge *et al.*, 2016).

In plants, metabolomics has been applied to characterize genetically modified varieties and to identify responses related to biotic and abiotic stresses (Carreno-Quintero *et al.*, 2013). In plant-pathogen interaction studies, metabolomics can unravel pathways hijacked by the pathogen and to predict resistance mechanisms (Allwood *et al.*, 2008; López-Gresa *et al.*, 2010). In sugarcane, metabolomics has been used to determine profiles related to sucrose accumulation (Bosch *et al.*, 2003; Glassop *et al.*, 2007), to evaluate metabolic fingerprints of genotypes with different susceptibility degrees to orange rust disease (Leme *et al.*, 2014), to distinguish between embryogenic and non-embryogenic callus tissue (Mahmud *et al.*, 2015) and to explore potential coproducts besides sucrose (Coutinho *et al.*, 2016).

The present study aims to determine changes in sugarcane metabolome in response to *S. scitamineum* colonization throughout disease progression. We used a combination of gas chromatography coupled to mass spectrometry (GC-TOF-MS) and liquid chromatography coupled to electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) complementary approaches. The results suggest a reprogramming in plant metabolism very early in response to *S. scitamineum* colonization. Metabolomics data corroborated previous hypothesis built on transcriptomics data of the same interaction in an independent experiment, and brings new information that will help to understand the biological mechanisms involved in smut disease as well as important features that can be used in breeding programs.

5.2. Material and methods

5.2.1. Ethics Statement

S. scitamineum SSC39 teliospores were collected as described by Taniguti *et al.* (2015). Experiments were performed using the smut susceptible Brazilian commercial variety of sugarcane, 'RB925345'. The healthy buds used to conduct the experiments were obtained from the IAC (Instituto Agronômico de Campinas, Centro de Cana), Ribeirão Preto, São Paulo, Brazil. No special permits were necessary for the teliospores or cane collection.

5.2.2. Experimental design

S. scitamineum SSC39 teliospores (Taniguti *et al.*, 2015) with viability >95% were used to inoculate single budded sets of 7 month-old plants of the "RB925345" genotype, classified as susceptible to smut. Prior to inoculation, plants were surface disinfected (Taniguti *et al.*, 2015). Puncture method was used for inoculation (10^6 teliospores.mL^{-1} in saline solution - $NaCl_2$ 0.85M). Mock-inoculated plants were prepared with saline solution (control plants). Inoculated and control plants were placed in greenhouse benches in a completely randomized experimental design. Samplings were made from buds 5 days after inoculation (DAI) and from the meristematic region of the main culm at 65 DAI, 100 DAI and 120 DAI, the last corresponds to the time immediately after whip emission. Each time point analyzed was represented by five biological replicates composed of pools of three plants for 5, 65 and 100 DAI samples. The 120 DAI replicates were represented by one plant (Figure 23). All samples were frozen in liquid nitrogen immediately after collection and stored at -80°C. Infected plants were compared to control samples of the same age.

5.2.3. Quantification of *S. scitamineum* DNA

Real-time qPCR was used to confirm and quantify *S. scitamineum* infection in each biological replicate. CTAB method was used for DNA extraction (Doyle and Doyle, 1990). qPCRs were made using as target the ribosomal Intergenic Spacer region (IGS) from *S. scitamineum* genome (Peters, 2016). Reactions consisted of 100 ng of total DNA, 0.2 μM of each primer, and 1× LuminoCt SYBR Green qPCR ReadyMix (Sigma-Aldrich), in a total volume of 12.5 μL. Cycling parameters were 95°C for 20 s, followed by 40 cycles of 95°C for 3 s and 60°C for 30 s. All reactions were run in an ABI 7500 Fast real-time PCR detection system (Applied Biosystems) in technical duplicates. Fluorescence (520 nm) was detected at the end of the elongation phase for each cycle. To evaluate amplification specificity, melt curve analysis was

performed at the end of each PCR run. The quantity of *S. scitamineum* DNA in each sample was determined by absolute quantification based on a standard curve obtained using DNA extracted from mixed cultures of *S. scitamineum* SSC39A and *S. scitamineum* SSC39B isolate. Quantifications were statistically analyzed using t-test (p-value < 0.05).

5.2.4. Metabolites extraction

Metabolites were extracted from 25 mg of grounded material. For GC-TOF-MS, extraction buffer was composed of methanol, chloroform and water (3:1:1, Gullberg *et al.*, 2004). The isotopically labeled succinic acid (D4, 98% - DLM 584-5), myristic acid (1, 2, 3 - $^{13}C_3$, 99% - CLM 3665 - 0.5) and palmitic acid (1, 2, 3, 4 - $^{13}C_4$) were used as internal standards. For LC-ESI-MS analysis, extraction buffer was composed by 99.875% of methanol and 0.125% of formic acid (De Vos *et al.*, 2007), containing 50 pmol of quercetin as internal standard.

Metabolites extraction followed (De Vos *et al.*, 2007) protocol with minor modifications. Initially, 0.5 mL of cold extraction buffer was added in each sample along with tungsten magnetic beads, and subjected to agitation in Vibration Mill (Retsch) for 30 seconds and 20 Hz. Beads were removed and samples sonicated for 15 min at 4°C and centrifuged for 10 min at 4°C, 16000 g. The supernatant was filtered (Millex 0.22 μM filter, Millipore) and stored at -80°C.

5.2.5. GC-TOF-MS and data processing

The organic phase (50 μL) was dried and derivatized as described in (Roessner *et al.*, 2001). 1 μl of the derivatized samples were analyzed on a Combi-PAL autosampler (Agilent Technologies GmbH, Waldbronn, Germany) coupled to an Agilent 7890 gas chromatograph coupled to a Leco Pegasus 2 time-of-flight mass spectrometer (LECO, St. Joseph, MI, USA) in split (1:40) and splitless mode described in Weckwerth *et al.* (2004).

Processing of the GC-TOF-MS data was performed in two stages. Initially chromatograms were exported to Leco ChromaTOF software (version 3.25) and performed the correction of baseline and data conversion to Network Common Data Form (NetCDF) format using the MassLynx software (Waters Corp, Milford, MA). Peak detection, retention time alignment, and library matching were performed using Target Search R-package (Cuadros-Inostroza et al., 2009) in R (R Core Team, 2015). Metabolites were quantified by the peak intensity of a selective mass. Metabolites intensities were normalized by dividing the fresh weight of each biological replicate, followed by the sum of total ion count (*TIC*) and Log_2 transformed. Metabolite data were normalized by dividing each raw value by the median of all measurements

of the experiment for one metabolite. The significance of metabolites was tested by t-test (p-value < 0.05).

5.2.6. LC-ESI-MS/MS

Metabolites from sugarcane meristem were analyzed in a mass spectrometer Q-TOF Ultima-API (Waters®), with ESI ionization source (Electrospray Ionization), coupled to an Acquity UPLC (Waters®). For UPLC chromatographic separation 5 μL of sample was injected in a reversed phase column (C18 100 x 2.1 mm 1.7 μM Acquity - Waters®). Two eluents were used as mobile phase: A (100% water containing 0.1% formic acid) and B (100% acetonitrile containing 0.1% formic acid). The mobile phase gradient used was: 95% A and 5% B for 6 minutes, 25% A and 75% B for 6 minutes, 5% A and 95% B for 1 minute. The capillary voltage was 3 kV and cone voltage of 35 kV. The temperature in the ionization source was 150°C and the desolvation temperature was 450°C. The nitrogen flow was 50 $L.h^{-1}$ in the cone and 550 $L.h^{-1}$ at the source. Data were acquired in positive and negative mode and centroid acquisition, in the mass dynamic range of m/z 100-1000, using MassLynx 4.1 software (Waters®).

The raw data was processed in MarkerLynx v 4.1 (Waters®) for alignment, noise removal, deconvolution, normalization using the quercetin *TIC* (*Total Ion Counts*) and obtaining the intensity of each possible metabolite, using 250 intensity as lower limit (threshold). Then the intensity of each metabolite was normalized by fresh weight (mg) of the corresponding sample.

Data processing was performed in Metaboanalyst 3.0 software (Xia *et al.*, 2015). First, data were filtered using interquartile range, log transformed and normalized using pareto scaling. Partial least squares discriminant analysis (PLS-DA) was applied to the metabolite dataset from the two subject classes (control and infected plants) of each time analyzed. PLS-DA was validated by leave-one out cross-validation. MetaboAnalyst 3.0 software was also used to generate Variable Importance on Projection (VIP) scores, which identify the best variables for discriminating between subject classes. Differential accumulation of metabolites was determined using t-test between infected and control samples of the same age. P-values were corrected using the Benjamini-Hochberg method and a cut-off of FDR < 0.05 was applied. Top 10 VIPs released from LC-ESI-MS analysis of each comparison were selected for fragmentation. LC-ESI-MS/MS analyzes were conducted on the same ionization conditions described above and fragmentation was performed using collision energy values ranging from 10 eV to 40 eV. Fragmentation data was compared to Metlin database (https://metlin.scripps.edu/index.php) to find possible metabolites, and ACD/MS Structure ID suite software was used to compare fragmentation profiles to theoretically fragmented metabolites from database and manually checked.

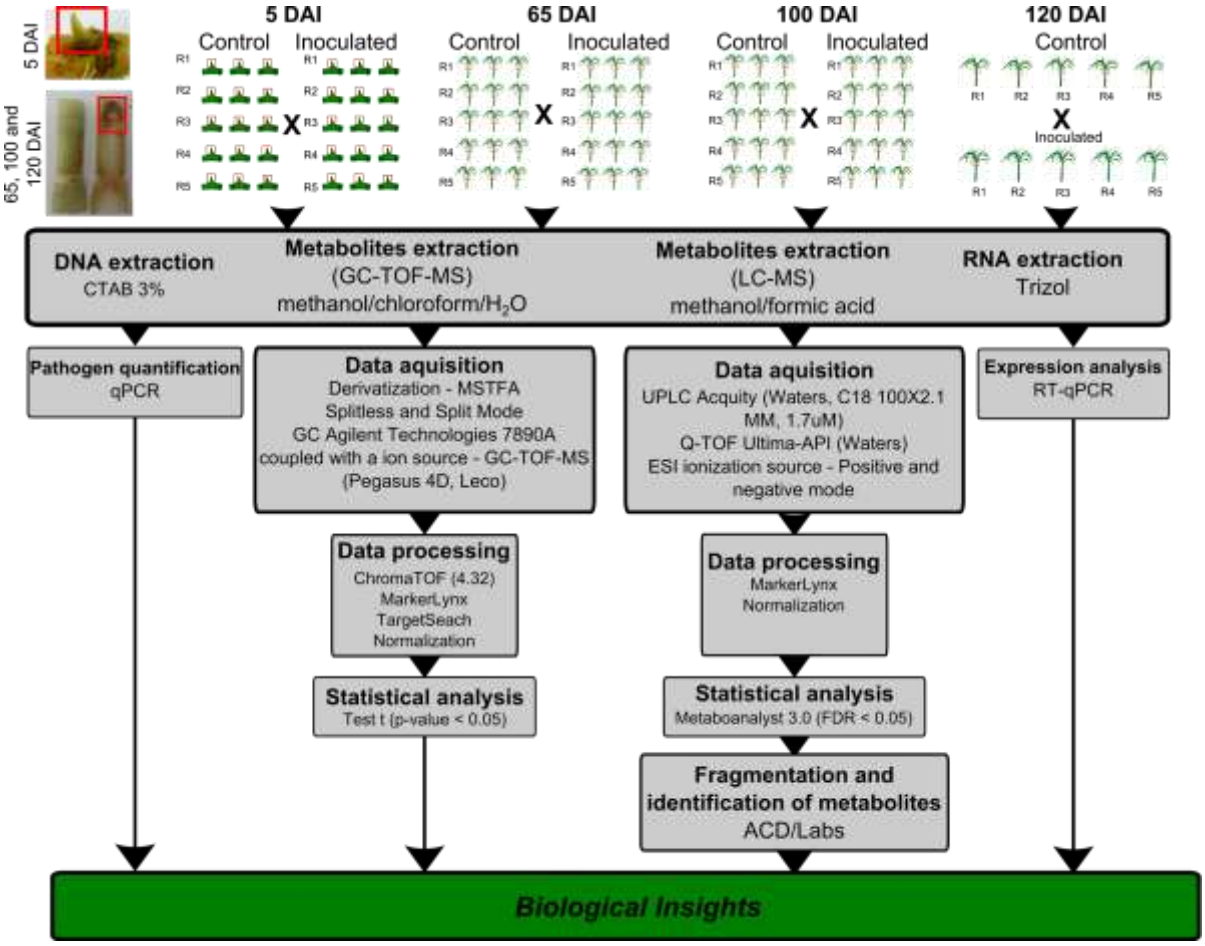


Figure 23. Schematic representation of experimental design, data acquisition and analysis to assess metabolic sugarcane responses to smut disease.

5.2.7. RT-qPCR

To further investigate biological evidences for metabolites accumulation, sugarcane genes were used as target to RT-qPCR reactions (Table 9). Primers were manually designed and quality verified using Gene Runner (http://www.generunner.net/) and NetPrimer (http://www.premierbiosoft.com/netprimer/) softwares. All RT-qPCRs were conducted in the 7500 Fast Real-Time PCR System (Applied Biosystems) using GoTaq® One-Step RT-qPCR System Kit (Promega). A reaction mixture containing 50 ng of RNA, 1 X of GoTaq® qPCRMaster Mix, 0.2 µM of each primer, 1 X µL of GoScript™ RT Mix and nuclease-free water to a final volume of 12.5 µL was prepared. Five biological replicates and two technical replicates were used. Cycling conditions were as follows: 37 °C for 15 min, 95 °C for 10 min, 40 cycles of 95 °C for 10 s, 60 °C for 30 s and 72 °C for 30 s. Primer specificity was confirmed obtaining the dissociation curve for every reaction. Sugarcane housekeeping genes encoding for polyubiquitin (Papini-Terzi, 2005) and GAPDH (Iskandar *et al.*, 2004) were used to normalize expression signals. PCR efficiencies and Cq values were obtained using the LinReg PCR program (Ramakers

et al., 2003). Relative changes in gene expression ratios were calculated by REST software (Pfaffl *et al.*, 2002). Control samples (mock-inoculated plants) were used as calibrators. The t-test was used to estimate significant changes in relative expression levels (p-value <0.05).

Table 9. Primers used to determine gene expression in RT-qPCR analysis.

Gene	Sequence	Reference
Acid invertase	5' GGAGGACGAGACCACACTC 3' 3' CGTTGTTGAAGAGGAACAC 5'	Schaker et al., 2016
Starch synthase	5' CCACGAAACACCATGATAGC 3' 3' GGCAATTCCTCCTCAGACAT 5'	This work
Xylan 1,4-beta-xylosidase	5'CCACGATTCTTGAGATTTCCTG 3' 3' GCCATCATCAGTCCAGTAATGC 5'	This work
Cellulose synthase	5' ATGGCTGATGGCACTCCTTG 3' 5' GTGATGCTGGAAACCTGGTCTC 3'	This work
Sucrose synthase	5'-TCCATCTACTTCCCCTTCACACAG-3' 3'-CTTCACCTTGTCCAGCCTTGC-5'	This work
GAPDH	5' CACGGCCACTGGAAGCA 3' 3' TCCTCAGGGTTCCTGATGCC 5'	Iskandar et al., 2004
Poliubiquitin	5' CCGGTCCTTTAAACCAACTCAGT 5' 3' CCCTCTGGTGTACTCCATTTG3'	Papinni-Terzi et al., 2005

5.2.8. Starch staining

Starch accumulation was evaluated in sugarcane infected plants after whip development in "R925345" genotype. Fresh cuts were made from whip region with intense sporulation, basis of whip, primary meristem and stem. Samples were stained using the potassium iodide-iodine reaction (I_2KI) for 5 min and observed under Light microscopy (Optika B-350 microscope). The images of I_2KI stained preparations were collected using an Optikam B5 digital camera.

5.2.9. Phenylalanine/Tyrosine ammonia-lyase phylogeny

The names and accession numbers of the sequences used in the phylogenetic analysis are presented in Appendix A. The amino acid multiple sequence alignments were created using ClustalW implemented in MEGA 6.06 Software (Tamura *et al.*, 2013) with default settings. The phylogenetic tree was built by MEGA 6.06 using a maximum likelihood algorithm, with the following settings: JTT model, 1000 replicates of bootstrap analyses, with best network interface (NNI) topology search.

5.3. RESULTS

5.3.1. Pathogen growth within plant tissues

The experimental design included analysis of the meristem region in four time points of sugarcane-smut interaction: two representing the limits of the colonization process (5 and 120

DAI) as studied before (Taniguti et al., 2015; Schaker et al., 2016), and two representing intermediate steps of the infection process (65 and 100 DAI). This design was used to explore further the molecular events described before at transcriptional and proteomics levels (Schaker et al., 2016; Barnabás et al., 2016) and narrow down time points that can reveal host candidate molecules potentially influencing fungal sporogenesis and consequently whip development.

qPCR technique using as template DNA extracted from the same samples of the metabolome analysis detected growing concentrations of the pathogen over time in inoculated plants (Figure 24). However, pathogen quantification among samples before (65 and 100 DAI) and after whip development (120 DAI) did not differ statistically (p-value > 0.05), indicating that pathogen index *per se* does not determine whip emission. These quantities represented less than 3 % of total DNA. Since we cannot discriminate metabolites from either plant or pathogen, the low abundance of pathogen DNA in the samples allowed us to postulate that the changes observed in metabolites abundance were mostly of plant origin.

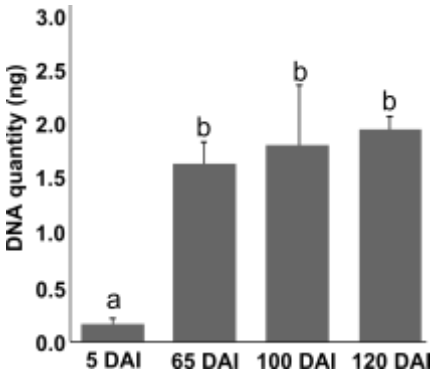


Figure 24. *S. scitamineum* DNA quantification using qPCR assay in samples used for metabolome analysis. Quantities are relative to 100 ng of total DNA. t-test (p-value < 0.05).

5.3.2. GC-TOF-MS general results

GC-TOF-MS approach is recognized as effective to determine changes in volatile compounds abundance or those able to be volatilized after derivatization, such as organic acids, amino acids, mono- and di-saccharides and others. Using this technique 73 metabolites were identified in sugarcane primary meristem (Appendix B). Each time point analyzed presented a particular set of compounds quantitatively altered by the smut infection compared to healthy plants of same age (p-value < 0.05) (Figure 25, Appendix B). The overall picture reflects that throughout time carbon partitioning is largely affected during fungal colonization and whip development disturbing the normal source-sink dynamics. Additionally, a shift in plant

meristematic metabolism occurs in between 65 and 100 DAI, and can be interpreted as a premature provision to whip emission.

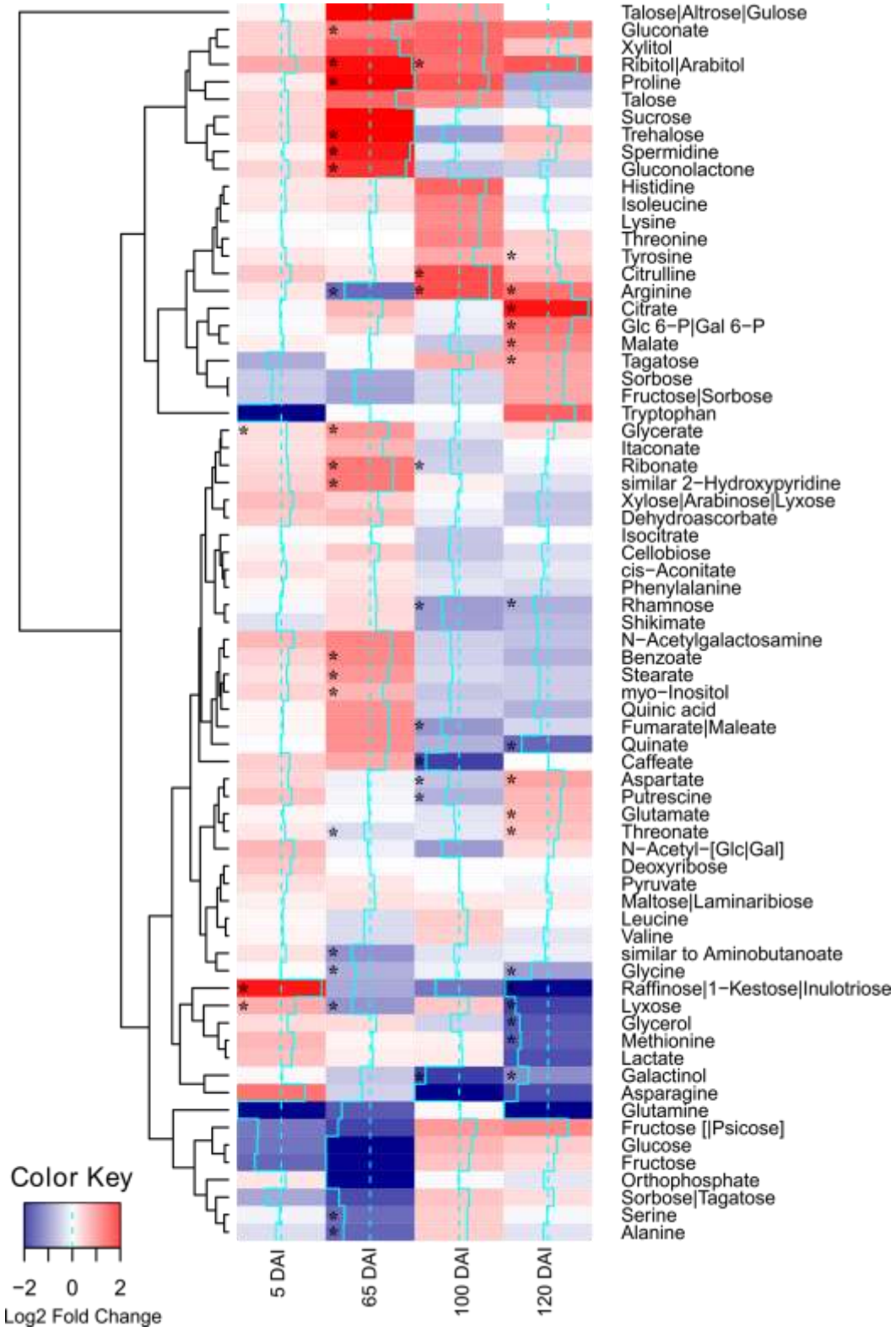


Figure 25. Heatmap of sugarcane metabolites identified in GC-TOF-MS analysis. Heatmap was build using Log2 Fold Change (Inoculated/Control) of relativized medians using gplots (Gregory et al., 2015) for R (R Core Team, 2015). Blue scale indicates low concentration in infected samples, and red scale indicates high concentration in infected samples. Squares "*" marked represent metabolites showing statistical significance in the comparison infected versus control (t-test, p-value < 0.05).

5.3.3. Early response to smut colonization

Soon after inoculation (5 DAI) infected plants accumulate significative (p-value < 0.05) higher levels of glycerate, lyxose and raffinose/1-kestose/inulotriose. Raffinose levels were approximately 10 times higher in infected samples, indicating an important role of this sugar during pathogen recognition. On the contrary, raffinose levels strongly decreased towards the sporogenesis and whip development.

5.3.4. Amino acids differential accumulation

Changes in amino acids accumulation were remarkable in smut infected plants (Figure 26), and more accentuated at 65 DAI, when the amount of proline (Pro) was increased and serine (Ser), alanine (Ala), arginine (Arg) and glycine (Gly) were reduced in infected plants (p-value < 0.05).

Before whip development, at 100 DAI, Arg and aspartate (Asp) were differentially accumulated in infected plants. Both amino acids were also affected after whip emission (120 DAI), i. e. Arg is kept significantly increased in infected samples, along with glutamate (Glu) and Gly, while Asp levels were reduced. Whip emission was also characterized by increased tyrosine (Tyr), which may be related to phenylpropanoid pathway, along with lower levels of methionine (Met), suggesting its catabolism toward ethylene synthesis, as previously suggested in transcriptomic analysis of the same interaction (Schaker *et al.*, 2016).

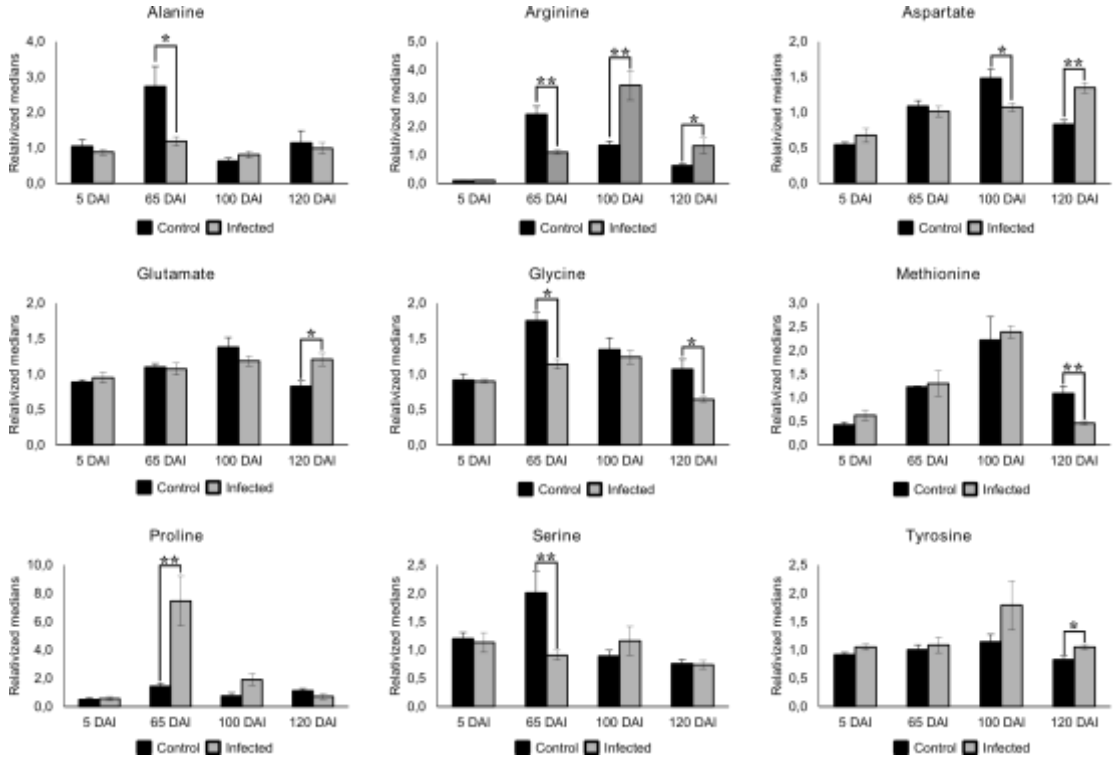


Figure 26. Amino Acids with significative levels alteration (p-value < 0.05) in sugarcane smut infected plants during disease progression. Relativized medians of A) alanine - Ala, B) arginine - Arg C) aspartate - Asp, D) glutamate - Glu, E) glycine - Gly, F) methionine - Met; G) proline - Pro, H) serine - Ser and I) tyrosine - Tyr in control and infected samples. "*" and "**" represent significative (T-test, p-value < 0.05 and p-value < 0.01, respectively) increase or reduction in metabolite level. Bars represent standard errors.

5.3.5. Carbon partitioning

Increased energetic requirements of infected plants was evidenced by the accumulation of intermediates of glycolysis, tricarboxylic acid cycle (TCA) and pentose phosphate pathway (PPP) even before whip development. This metabolic response can bring on negative impacts in stem sucrose accumulation, even if whip id not formed or delayed. After whip emission, significative increased levels of glucose 6P, malate and citrate were detected in infected plants (Figure 25). Changes in carbon flux toward respiratory pathways may be connected to increased gene expression of soluble acid invertase detected in 100 and 120 DAI samples (Figure 27). However, sucrose in meristems of infected samples was not altered, probably because they are not the main sucrose-storage tissues. Over-expression of starch synthase before whip emission (100 DAI) (Figure 27) may indicates starch accumulation as a drain of carbon in infected samples (Figure 28).

Differential accumulation of other carbohydrates in infected samples may contribute to sugar signaling and differential carbon allocation. For instance, trehalose levels had a subtle increase in infected samples at 65 DAI, and higher levels of ribitol/arabitol were found in

infected samples at 65 and 100 DAI. Additionally, arabitol is synthesized by reduction of either arabinose or lyxose, which may explain the significative low levels of lyxose in infected plants at 65 DAI. After whip emission, raffinose, a metabolite involved in several aspects of plants metabolism and signalization, and also positively correlated to sucrose accumulation in sugarcane (Glassop *et al.*, 2007), was significantly reduced in infected plants, together with its precursor galactinol, which was reduced both 100 and 120 DAI samples.

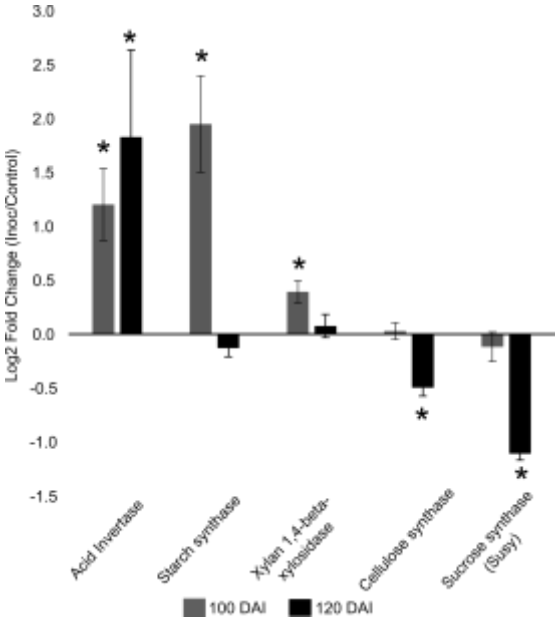


Figure 27. RT-qPCR analysis of key sugarcane genes related to smut disease. "*": indicates significative reduction or increased levels of transcripts in infected plants compared to control ones (t-test, p-value 0.05).

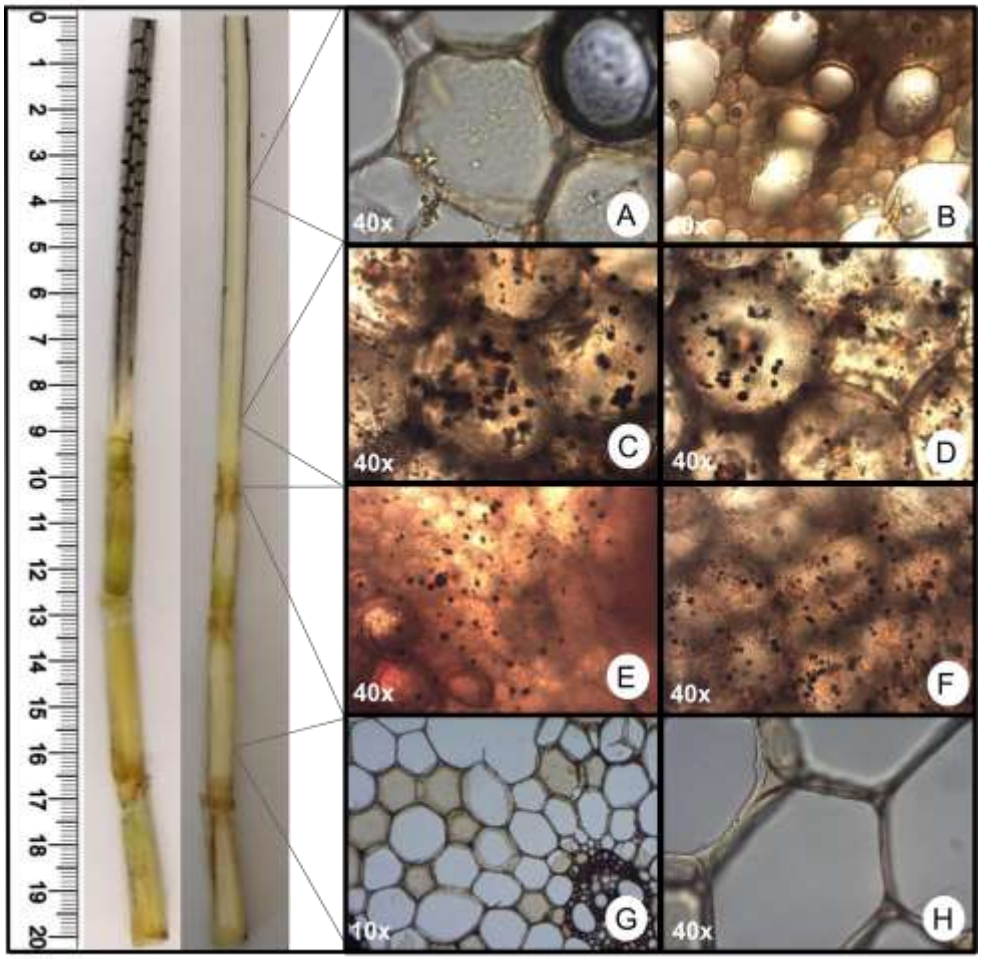


Figure 28. Accumulation of starch in smut infected sugarcane. Potassium iodide-iodine reaction (I_2KI) staining was used to detect starch by light microscopy in different sections of sugarcane infected samples after whip development. A-B) Whip region with intense sporulation. C-D) Basis of whip. E-F) Primary meristem. G-H) Stem.

5.3.6. Cell wall precursors

Several metabolites related to cell wall biogenesis were identified in GC-TOF-MS analysis, including shikimate, phenylalanine, tyrosine, xylose, rhamnose, cellobiose and caffeate. All of them presented a similar pattern of regulation in late stages of the smut disease - 100 and 120 DAI (Figure 25). Significative reduction was detected for rhamnose and caffeate at 100 DAI, and rhamnose after whip emission, suggesting that a weakening of cell wall may allow whip growth.

The gene expression analysis (RT-qPCR) of the following sugarcane genes: xylan 1,4-beta-xylosidase, cellulose synthase and sucrose synthase, was performed to further investigate plant responses related to whip development and cell wall constitution (Figure 27). Cell wall weakening may be related with increased expression of hemicellulose degrading xylan 1,4-beta-xylosidase early before whip development. Cellulose synthase and sucrose synthase expression

did not change before whip but had a significative reduction (p-value < 0.05) at 120 DAI - after whip.

Higher levels of tyrosine in infected samples suggested that phenylpropanoid pathway is affected as previously detected (Barnabas *et al.*, 2016). Even though several studies inferred responses of PAL (phenylalanine ammonia-lyase) in smut infected plants, changes in Phe levels were not observed in our work. Considering that in grasses part of lignin is synthesized from tyrosine by PTAL, a bifunctional phenylalanine-tyrosine ammonia-lyase, we analyzed the amino acid sequence of sugarcane PAL/PTAL, and found that those genes responsive to smut in transcriptional analysis and at protein levels encoded protein from PTAL family (Figure 29), indicating that tyrosine accumulation is indeed related to increased levels of lignin in smut infected samples.

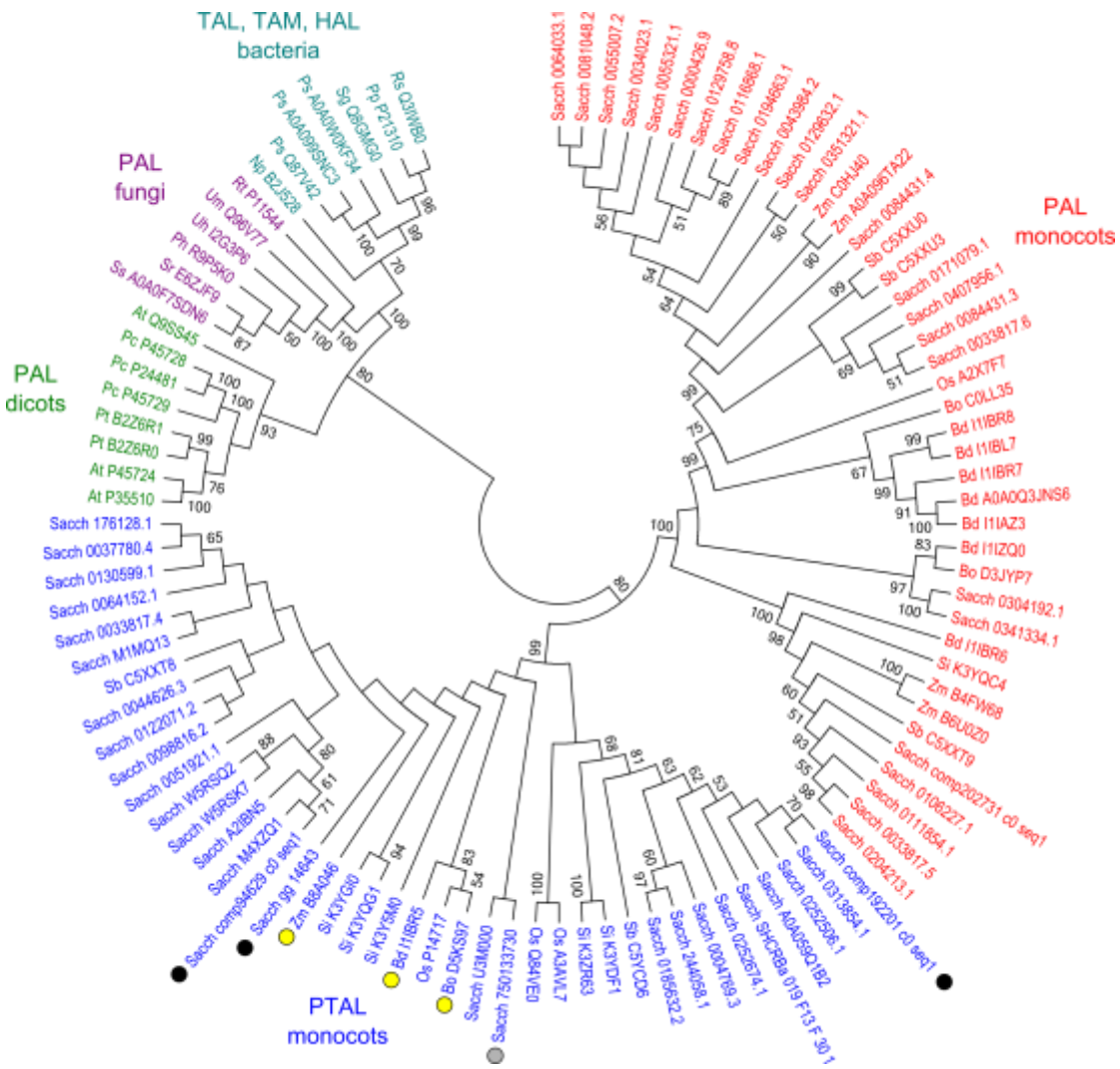


Figure 29. Phylogenetic tree of PTAL and PAL in plants and fungi, and TAL, tyrosine ammonia-mutase (TAM) and histidine ammonia-lyase (HAL) in bacteria. Protein sequences were obtained from Uniprot. Black circles: sugarcane genes up-regulated in smut infected plants after whip development (Schaker *et al.*, 2016); Gray circle: sugarcane protein identified exclusively in sugarcane smut-infected plants after whip development (Barnabas *et al.*, 2016); Yellow circles: previously reported PTAL proteins.

5.3.7. LC-ESI-MS general results

Untargeted LC-ESI-MS was implemented as a complementary approach to describe changes in sugarcane secondary metabolism during the progression of smut disease. Ionization in positive mode detected 254, 216, 262 and 260 non-redundant *m/z* in samples from 5, 65, 100 and 120 DAI, respectively (Table 10), while negative ionization resulted in 290, 223, 235 and 232 non-redundant *m/z* detected for the same samples (Table 10). PLS-DA plots (Figure 30) showed the discrimination between control and infected plants, considering 95% of confidence. In 5 DAI samples, 33 and 83 metabolites were statistically quantitatively altered by the infection (FDR ≤ 0.05) using positive and negative ionization, respectively. These numbers increased toward disease progression (Table 10), indicating that changes of complex molecules composition after whip emission were relevant and still poorly understood.

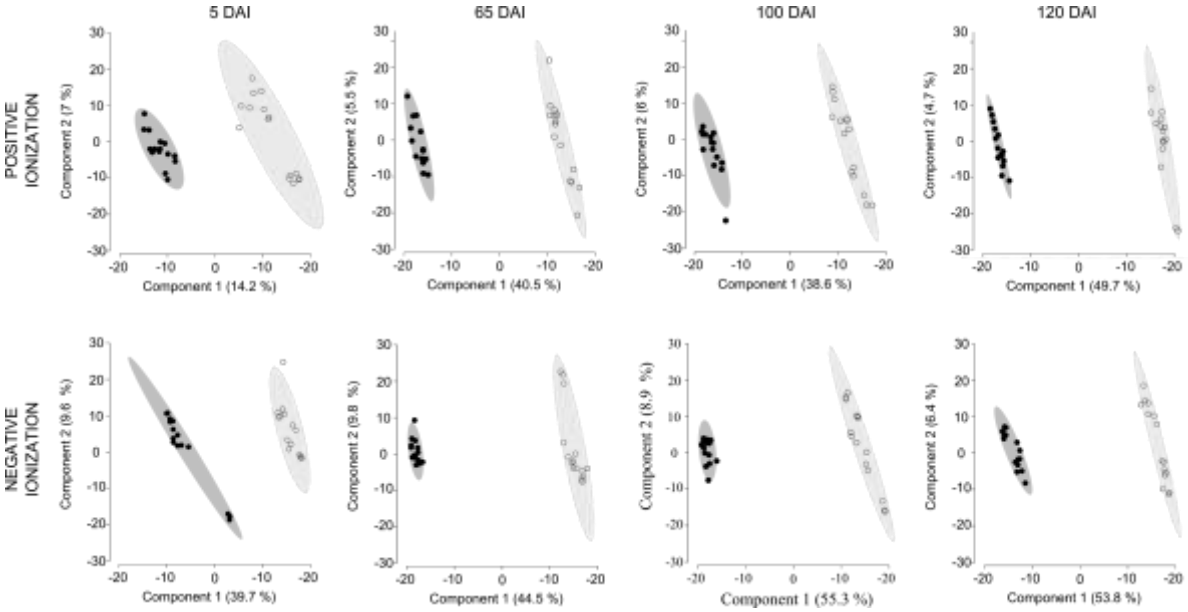


Figure 30. PLS-DA plots of LC-ESI-MS metabolome profile using positive and negative ionization in 5, 65, 100 and 120 DAI samples. Black dots represent biological and technical replicates of control plants, empty dots represent infected plants. Ellipses indicates the 95% confidence region.

Table 10. Number of *m/z* detected by LC-ESI-MS analysis and number those differentially accumulated (FDR ≤ 0.05) comparing inoculated and control samples of the same age.

Sample	Positive Ionization		Negative Ionization	
	Total number of m/z detected	m/z regulated (FDR < 0.05)	Total number of m/z detected	m/z regulated (FDR < 0.05)
5 DAI	254	33	290	83
65 DAI	216	107	223	87
100 DAI	262	113	235	137
120 DAI	260	154	232	173

5.3.8. Sugarcane-smut secondary metabolism profile

Using the set of differentially accumulated *m/z* (FDR < 0.05) detected in LC-ESI-MS approach, four-way Venn diagrams were obtained and showed four significative molecules (*m/z*) shared among all time points analyzed of the sugarcane-smut interaction (Figure 31 A-C). MS-MS fragmentation allowed the identification of two of these metabolites (Appendix C): *m/z* 475.1191, with a fragmentation pattern of apigenin 7-O-(6"-O-acetylglucoside), and *m/z* 445.1002 identified as 3'-O-methylderhamnosylmaysin. Apigenin 7-O-(6"-O-acetylglucoside) synthesis was suppressed in infected plants during disease progression, and increased after whip development, becoming an important marker metabolite related to health sugarcane plants.

VIP scores (Figure 31D) were used to identify metabolites that most contributed to discriminate control and inoculated plants of the same age. The top 10 VIPs from each comparison were fragmented, leading to identification of 25 metabolites from positive ionization and 6 from negative ionization (Appendix C and D). One metabolite was detected in both ionization methods among the VIPs (Figure 32).

In early infection (5 DAI), pratensin B, ax-4'-hydroxy-3'-methoxymaysin, quercetin 3-(2-caffeoylsophoroside) 7-glucoside and a metabolite with fragmentation pattern similar to the antifungal compound *Sch59884* were detected only in infected plants, indicating activation of defense responses. Other metabolites were detected only in control samples in early infection, for example 5-oxoavermectin, 2-methylhexanoyl-CoA and 8-azaadenosine. Similarly, metabolites related to maysin, such as 3'-O-methylderhamnosylmaysin and derhamnosylmaysin were detected control samples at 5, 65 and 100 and 5, 100 and 120 DAI days of plant growth, respectively (Figure 32).

Metabolites corresponding to *m/z* 392.1120, 432.1012, 410.1267 and 409.1102 were detected only in infected plants at 65, 100 and 120 DAI. One of them had a fragmentation pattern similar to Fusarin C, a fungal origin metabolite classified as a mycotoxin. Two other identified metabolites (*m/z* 671.1332 - R-Skyrin 2-xyloside, and *m/z* 453.1112 - urdamycione B) were detected only in control samples at 65, 100 and 120 days of plant growth. After whip emission, other metabolites were detected only in control samples, including glycero-3-phospho-(1'-myo-inositol), heliocide and primflaside (Figure 32).

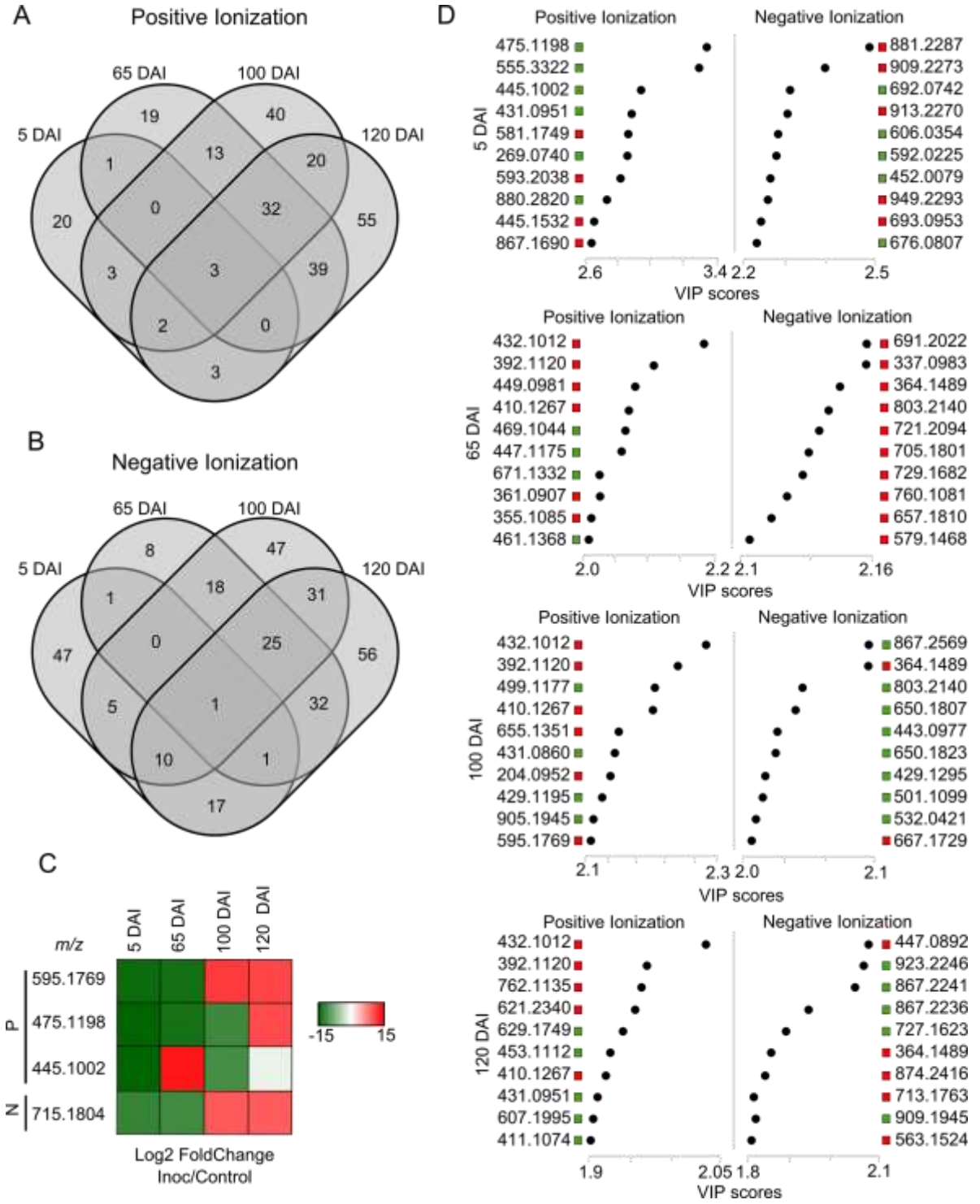


Figure 31. LC-ESI-MS metabolomics analysis in sugarcane-smut interaction. Four-way Venn diagram representing differentially accumulated (FDR ≤ 0.05) non-redundant *m/z* of each time point analyzed among infected and control plants of the same age using (A) positive and (B) negative ionizations. C) Heatmap shows the dynamics of these metabolites during the progression of sugarcane smut disease. It was obtained with gplots (Gregory et al., 2015) for R (R Core Team, 2015) with Log2 Fold Change (inoculated/control) values of shared differentially accumulated *m/z* from Venn diagrams. "P": Positive ionization, "N": negative ionization. D) Top 10 *m/z* based on VIP scores from PLS-DA for each time point analyzed in positive and negative ionization. These *m/z* were submitted to LC-ESI-MS/MS analysis to confirm their identity. The x-axis shows the correlation scores and y-axis corresponds to LC-ESI-MS *m/z*. Regulation of each *m/z* in inoculated plants compared to control is represented as colored squares. Red squares represent those were exclusively identified in inoculated samples, green squares represent *m/z* that were exclusively identified in control samples.

	m/z	Metabolite	Metlin ID	ΔMass (Da)	Retention time (min)	Ionization	5 DAI	65 DAI	100 DAI	120 DAI
Positive Ionization	555.3322	6,8a-Seco-6,8a-deoxy-5-avermectin-"1b" aglycone	63702	0.008	10.1	ESI(+)				
	880.282	2-Methylhexanoyl-CoA	62405	0.078	8.64	ESI(+)				
	269.074	8-Azaadenosine	68949	0.018	4.77	ESI(+)				
	593.2038	ax-4"-Hydroxy-3-'-methoxymaysin	86824	0.025	5.35	ESI(+)				
	867.169	Sch 59884	70339	0.164	6.24	ESI(+)				
	445.1532	Pratensin B	51716	0.011	7.18	ESI(+)				
	431.0951	Derhamnosylmaysin	48939	0.005	5.69	ESI(+)				
	355.1085	3',4'-Dihydrooxepino-6'-hydroxybutein	52073	0.002	1.54	ESI(+)				
	361.0907	2,3-Dihydro-2-(4-hydroxyphenyl)-5,6,7,8-tetramethoxy-4H-1-benzopyran-4-one	70442	0.030	2.85	ESI(+)				
	655.1351	Herbacetin 7-methyl ether 3-(2"-(E)-feruloyglucoside	51554	0.023	6.03	ESI(+)				
	629.1749	Pinocembrin 7-O-neohesperidoside 3"-O-acetate	52629	0.080	4.79	ESI(+Na)				
	429.1195	1-dodecanoyl-glycero-3-phospho-(1'-sn-glycerol)	79999	0.098	5.01	ESI(+)				
	599.1284	Dexamethazone metasulfobenzoate sodium	69517	0.036	5.53	ESI(+)				
	437.1302	Chapelieric acid methyl ester	48441	0.059	6.96	ESI(+)				
	607.1995	1-(9Z-heptadecenoyl)-glycero-3-phospho-(1'-myo-inositol)	81197	0.003	4.79	ESI(+Na)				
	411.1074	Heliocide H3	87167	0.102	5.01	ESI(+)				
	671.1332	(R)-Skyrin 2-xyloside	91464	0.001	5.32	ESI(+)				
	469.1044	Artoindonesianin B	49959	0.074	5.03	ESI(+)				
	453.1112	Urdamycinone B	63744	0.036	6.07	ESI(+)				
	392.112	Proacaciberin	66916	0.036	5.31	ESI(+)				
	432.1012	Fusarin C	72964	0.093	5.31	ESI(+)				
	410.1267	Acidissiminol epoxide	95258	0.099	5.32	ESI(+)				
	409.1102	6,8-Dihydroxy-1,7-diprenylxanthone-2-carboxylic acid	92248	0.047	2.64	ESI(+)				
	475.1198	Apigenin 7-O-(6'-O-acetylglucoside)	48805	0.004	2.05	ESI(+)				
	445.1002	3'-O-Methylderhamnosylmaysin	49185	0.006	6.36	ESI(+)				
Negative Ionization	949.2293	Quercetin 3-(2-caffeoylsophoroside) 7-glucoside	95300	0.004	5.47	ESI(-)				
	337.0983	Paratocarpin K	52798	0.017	2.85	ESI(-)				
	691.2022	Curcumin diglucoside	64200	0.029	2.03	ESI(-)				
	443.0977	3'-O-Methylderhamnosylmaysin	49185	0.008	6.36	ESI(-)				
	667.1729	Okanin 4-methyl ether 4'-O-(2"-O-caffeoyl-6"-O-acetylglucoside)	51983	0.001	6.33	ESI(-)				
	727.1623	Primflaside	50621	0.018	1.53	ESI(-)				

Figure 32. Metabolites identified in LC-ESI-MS/MS analysis using ACD/Labs software to theoretical fragmentation of structures from Metlin database (https://metlin.scripps.edu/index.php). Green filled squares: metabolites detected only in control samples; red filled squares: metabolites identified only in infected samples. Fragmentation patterns are presented in Appendix C and D.

5.4. Discussion

Metabolomics is recognized as a powerful tool to describe plant responses to several stimuli. Integration of GC-TOF-MS and LC-ESI-MS/MS complementary tools allowed us to identify a set of metabolites involved in several aspects of plant growth and signalization with altered levels in meristem of smut infected samples. These metabolic responses initiated since shortly after inoculation and continued after whip emission.

Five days after inoculation this compatible interaction was characterized by the presence of compounds structurally similar to those with antifungal activities, and by an increased level of raffinose. Raffinose is synthesized from conjugation of galactinol and sucrose (Sengupta *et al.*, 2015). The enzyme galactinol synthase (GolS; EC 2.4.1.123) is a key component of the galactinol production and its overexpression is correlated to increased resistance to pathogens (Kim *et al.*, 2008). External application of galactinol in tobacco leads to expression of genes encoding PR1a, PR1b, and NtACS1, which are well-known defense-related proteins (Kim *et al.*, 2008).

Additionally, soluble sugars such 1-kestose and raffinose accumulate system-wide in non-infected plant parts, suggesting their function as transportable stress signals in biotic stresses (Moghaddam and Van Den Ende, 2012).

Our previous sugarcane-smut transcriptomic data on the same interaction did not detected changes of *golS* gene expression but instead revealed the repression of a α-galactosidase gene (EC 3.2.1.22), involved in raffinose breakdown (Schaker *et al.*, 2016), which may be the origin of raffinose accumulation in infected plants. Studies on protein differential accumulation in early smut infection also detected reduced levels of α-galactosidase in both resistant and susceptible genotypes, indicating that it is a general response to smut (Su *et al.*, 2016). Raffinose accumulation has been associated with response to oxidative burst, which may function as scavenger of ROS (van der Hoorn and Kamoun, 2008; Sengupta *et al.*, 2015), and as a compound able to stabilize cell membrane in unfavorable conditions (Hincha *et al.*, 2003). It is known that sugarcane promotes oxidative burst in response to smut infection (LaO *et al.*, 2008; Su *et al.*, 2014). Recently, details of ROS metabolism revealed that sugarcane resistant genotypes maintain high levels of oxygen peroxide by a SOD independent pathway potentially to activate defense responses (Peters, 2016). Raffinose accumulation is an attractive candidate strategy to act in conjunction with the antioxidant system in an attempt to restrain plant cell damage due to the oxidative burst in response to pathogen attack. Also, raffinose is usually less accumulated in meristematic regions; the most abundant sugars in these regions are glucose and fructose (Glassop *et al.*, 2007). Raffinose is mostly detected in more mature tissues and is positively associated to sucrose levels. Interestingly, raffinose-related biosynthesis pathway was inhibited 100 DAI and after whip development, with coherent results obtained in both metabolomics and transcriptomic analysis (Figure 33).

Intermediates of the carbon central metabolism composed the major group of regulated metabolites all over smut disease progression. Sugarcane meristems are actively growing and accumulate higher levels of amino acids and metabolites associated with TCA cycle compared to stem tissues (Glassop *et al.*, 2007). In smut-infected samples, it was detected even higher levels of compounds associated with glycolysis, TCA and PPP in the meristem. This result is also corroborated by RNAseq and proteomic data described for whipped sugarcane (Figure 33, Barnabas *et al.*, 2016; Schaker *et al.*, 2016). Up-regulation of energy-related pathways is a conserved response to demands of biotic stresses, such as the synthesis of secondary metabolites (Less *et al.*, 2011; Rojas *et al.*, 2014). Even in sugarcane genotypes resistant to smut, i. e. those that do not emit whip, *S. scitamineum* colonization is detected (Carvalho *et al.*, 2016) potentially interfering with carbon allocation. In addition, increased energetic metabolism network may also

imply that stems reduce sucrose accumulation, one of the smut susceptible symptoms yet not investigated in resistant genotypes colonized by smut.

Regarding carbon storage, it was detected an increased expression of starch synthase sugarcane gene at 100 DAI that we speculated that starch may be the source to feed the whip development in late moments of the interaction. RNAseq data from plants after whip development showed up-regulation of alpha-amylase, involved in starch breakdown (Schaker *et al.*, 2016).

Other relevant metabolic change in infected samples concerned amino acids accumulation, mainly 65 DAI. At this time point, infected samples accumulated proline and reduced levels of alanine, arginine, glycine and serine. Plants are known to increase proline levels during stress, acting as potent osmolyte, metal chelator, antioxidant, or signaling molecule associated with the hypersensitive response (Fabro *et al.*, 2004; Hayat *et al.*, 2012; Qamar *et al.*, 2015; Verslues and Sharma, 2010). On the other hand, proline accumulation may play a role in maintain a favorable environment for the fungal development in terms of protection against abiotic stresses such as UV light, heat, salt, and hydrogen peroxide (Chen and Dickman, 2005). Increased proline levels may be related to GABA (γ-Aminobutyric acid) reduced levels in infected plants 65 DAI. The GABA synthesis potentially compete for the same substrate (Verslues and Sharma, 2010).

Amino acids responsive to whip development, i.e. Tyr and Met, detected also in previous studies of differential gene expression (Figure 33, Schaker *et al.*, 2016) and differential protein accumulation (Barnabas *et al.*, 2016) supported the hypothesis stressed by several authors that increased lignin contents and ethylene imbalance are modulated by the fungus in this pathosystem. Lignin is a principal component of plant cell walls and was thought to be mostly produced from L-phenylalanine. However, in grasses nearly half of the plant's lignin is actually made through fewer steps via L-tyrosine, using a different path of that related to L-phenylalanine leading to formation of 4-coumarate (Barros *et al.*, 2016). Other plausible hypothesis relies on the fact that amino acids accumulation is regulated by feedback inhibition. Tyr accumulation in smut-infected plants may act as a positive regulator of Phe biosynthesis through activation of arogenate dehydratase (ADT), ensuring that the major carbon flux is directed toward Phe biosynthesis (Galili *et al.*, 2016; Maeda and Dudareva, 2012). Transcriptional and protein analysis also detected that after whip emission levels of phenylalanine ammonia lyase (PAL), cinnamoyl-CoA reductase, caffeic acid 3-O-methyltransferase, caffeoyl-CoA O-methyltransferase and peroxidases were increased (Schaker *et al.*, 2016), confirming changes in carbon allocation towards phenylpropanoids pathways. Plants in normal development can direct more than 30% of

photosynthetically fixed carbon through the vascular system to synthesize lignin via the phenylpropanoid pathway (Maeda and Dudareva, 2012). Caffeate levels were significantly reduced in infected samples 100 DAI returning to regular levels (control levels) after whip emission. This scenario suggests that whip emission is indeed a drain of lignin precursors.

Met is considered a fundamental metabolite of plant cells involved in the biosynthesis of ethylene and polyamines via SAM (S-adenosylmethionine) (Roje, 2006). This pathway is regulated by a feedback mechanism, where levels of the first committed enzyme of Met biosynthesis, cystathionine γ-synthase (CGS), is downregulated by SAM (Chiba *et al.*, 2003). Reduced levels of Met in infected samples after whip emission may be related to its conversion to SAM, which in turn negatively affects Met synthesis. The transcriptome data revealed that CGS is downregulated whereas S-adenosylmethionine synthetase is up-regulated in infected samples after whip development (Figure 33, Schaker *et al.*, 2016). Accordingly, the protein level of methionine synthase is down-regulated and SAM up-regulated at the same time (Barnabas *et al.*, 2016). Furthermore, increased expression of ACC oxidase indicates that SAM can be used as precursor to ethylene biosynthesis in infected samples, a plant hormone known to be positively correlated with activation of phenylpropanoids pathway (Ecker and Davis, 1987; Guo and Ecker, 2004).

Phenylpropanoid pathway is also the source of precursors involved in the secondary metabolism. In this study a surprising response was obtained. A compound related to maysin and its derivatives were identified in all time points analyzed by the LC-ESI-MS analysis. Maysin is recognized by its potential to confer resistance against the lepidopteran corn earworm in maize (Byrne *et al.*, 1996). Two metabolites precursors of its synthesis (derhamnosylmaysin and 3'-O-methylderhamnosylmaysin) were detected only in control samples in all time points, indicating that this pathway is suppressed in response to *S. scitamineum* colonization. It may be worth to evaluate if this same pattern is observed in other sugarcane genotypes and correlate them with plant resistance.

LC-ESI-MS approach also allowed the identification of a *S. scitamineum* metabolite similar to Fusarin C. This mycotoxin belongs to the class of acyl-tetramic acids, which is found in several fungus (Song *et al.*, 2004). The coding genomic region of Fusarin C in *Fusarium moniliforme* comprises a genomic fragment encoding a Type I PKS fused to a nonribosomal peptide synthase module (Song *et al.*, 2004). In *S. scitamineum*, studies on the genetic background related to mycotoxins biosynthesis is still missing. However, analysis of sugarcane juices revealed the presence of important mycotoxins, such as aflatoxin B1 and G1 (Abdallah *et al.*, 2016). These findings point to an important issue concerning food security, since sugarcane is the main source

of sugar, one of the most consumed foods worldwide, and hosts a large diversity of potentially toxin-producing fungi.

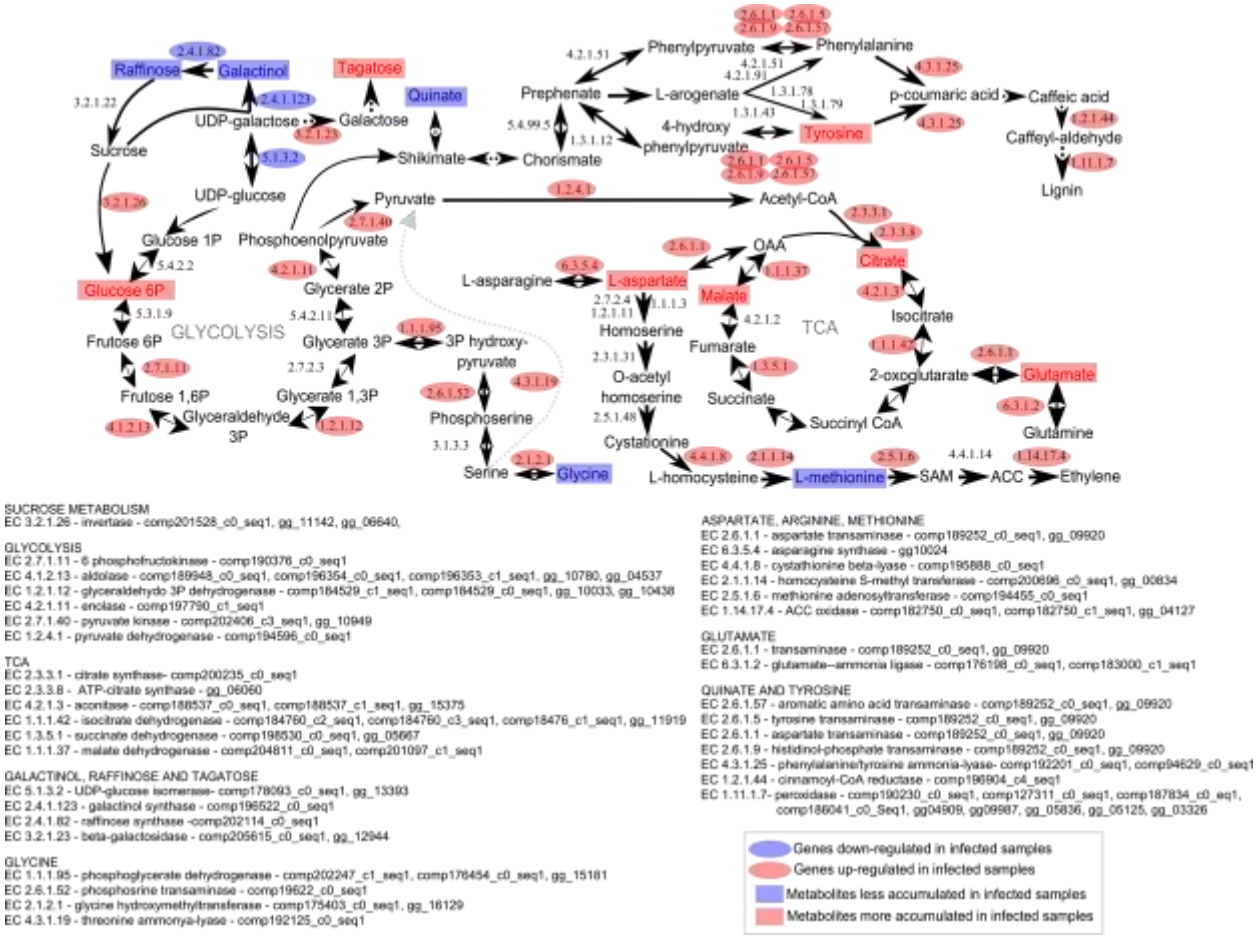


Figure 33. Metabolomic and transcriptomic responses in sugarcane related to whip emission 120 DAI of *S. scitamineum*. Transcriptomic data were obtained before using the same sugarcane genotype and experimental design (Schaker *et al.*, 2016).

5.5. Concluding remarks

This work presented an overall picture of the metabolites identified at the meristematic region of plants infected and non-infected with *S. scitamineum* in four time points of the smut disease progression, and stand as the first report involving metabolomics of a sugarcane pathosystem. The metabolome data was integrated to previously obtained transcriptomic analysis performed in independent experiments with the same sugarcane genotype and experimental design. This complementary information let to the proposal of the following working model for smut disease establishment and progression.

Since early infection meristematic cells respond to the pathogen colonization by accumulating raffinose, which may act directly toward ROS neutralization or as a signaling molecule. With disease progression, a shift in plant metabolism was detected, which may be

interpreted as a lead up to whip emission, implying in major changes in carbon allocation. For instance, gene expression analysis indicates that starch is the preferred way of storage in infected meristems, which may later be degraded to feed whip development, as suggested by the overexpression of an alpha-amylase gene after whip emission. Congruent results concerning metabolomic and transcriptional responses after whip emission were detected in several aspects of plant metabolism, including energetic, ethylene and cell wall pathways. The whip formation seems to relies on Tyr metabolism, suggesting the phenylpropanoid pathway as a carbon sink through the overexpression of a bifunctional PTAL. At the same time a weakening of cell wall may allow whip emission from meristem.

Compounds of the secondary metabolism reveals an active defense response that should be better studied in resistant varieties, and gave rise to metabolites which can be exploited as biomarkers, bringing new opportunities to study resistance mechanisms in sugarcane and improve breeding practices related to smut.

REFERENCES

Abdallah, M., Krska, R. and Sulyok, M. (2016) Mycotoxin Contamination in Sugarcane Grass and Juice: First Report on Detection of Multiple Mycotoxins and Exposure Assessment for Aflatoxins B1 and G1 in Humans. *Toxins (Basel).* **8**, 343.

Allwood, J.W., Ellis, D.I. and Goodacre, R. (2008) Metabolomic technologies and their application to the study of plants and plant–host interactions. *Physiol. Plant.* **132**, 117–135.

Barnabas, L., Ashwin, N.M.R., Kaverinathan, K., et al. (2016) Proteomic analysis of a compatible interaction between sugarcane and Sporisorium scitamineum. *Proteomics* **16**, 1111–22.

Barros, J., Serrani-yarce, J.C., Chen, F., Baxter, D., Venables, B.J. and Dixon, R.A. (2016) Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nat. Plants* **2**, 1–9.

Bindon, K.A. and Botha, F.C. (2002) Carbon allocation to the insoluble fraction, respiration and triose-phosphate cycling in the sugarcane culm. *Physiol. Plant.* **116**, 12–19.

Bolton, M.D. (2009) Primary Metabolism and Plant Defense—Fuel for the Fire. *Mol. Plant-Microbe Interact.* **22**, 487–497.

Borrás-Hidalgo, O., Thomma, B.P.H.J., Carmona, E., Borroto, C.J., Pujol, M., Arencibia, A. and Lopez, J. (2005) Identification of sugarcane genes induced in disease-resistant somaclones upon inoculation with Ustilago scitaminea or Bipolaris sacchari. *Plant Physiol. Biochem.* **43**, 1115–1121.

Bosch, S., Rohwer, J.M. and Botha, F.C. (2003) The sugar metabolome. *Proc S Afr Sug Technol Ass* **7**, 129–133.

Botha, A. and Whittaker, F.C. (1997) Carbon Partitioning during Sucrose Accumulation in Sugarcane Internodal Tissue. *Plant Physiol.* **115**, 1651–1659.

Botha, F.C. and Moore, P.H. (2014) Biomass and bioenergy. In Sugarcane - Physiology, biochemistry and functional biology. (Moore, P.H. and Botha, F.C., eds), pp. 521–541. Oxford, UK: Wiley-Blackwell.

Bu scher, J.M., Czernik, D., Ewald, J.C., Sauer, U. and Zamboni, N. (2009) Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. *Anal. Chem.* **81**, 2135–2143.

Byrne, P.F., Mcmullen, M.D., Snook, M.E., MUSKETt, T.A., THEURIt, J.M., Widstromt, N.W., Wisemani, B.R., Coe, E.H. and Phillips, R.L. (1996) Quantitative trait loci and metabolic pathways: Genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Agric. Sci.* **93**, 8820–8825.

Carreno-Quintero, N., Bouwmeester, H.J. and Keurentjes, J.J.B. (2013) Genetic analysis of metabolome–phenotype interactions: from model to crop species. *Trends Genet.* **29**, 41–50.

Carvalho, G., Quecine, M.C., Longatto, D.P., et al. (2016) *Sporisorium scitamineum* colonisation of sugarcane genotypes susceptible and resistant to smut revealed by GFP-tagged strains. *Ann. Appl. Biol.* **169**, 329–341.

Chandra, A. (2011) Physio-Biochemical and Molecular Approaches to Enhance Sucrose Content in Sugarcane: Indian Initiatives. *Sugar Tech* **13**, 315–321.

Chen, C. and Dickman, M.B. (2005) From The Cover: Proline suppresses apoptosis in the fungal pathogen Colletotrichum trifolii. *Proc. Natl. Acad. Sci.* **102**, 3459–3464.

Chiba, Y., Sakurai, R., Yoshino, M., Ominato, K., Ishikawa, M., Onouchi, H. and Naito, S. (2003) S-adenosyl-L-methionine is an effector in the posttranscriptional autoregulation of the cystathionine gamma-synthase gene in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 10225–30.

Coutinho, I.D., Baker, J.M., Ward, J.L., Beale, M.H., Creste, S. and Cavalheiro, A.J. (2016) Metabolite Profiling of Sugarcane Genotypes and Identification of Flavonoid Glycosides and Phenolic Acids. *J. Agric. Food Chem.* **64**, 4198–4206.

Doidy, J., Grace, E., Kühn, C., Simon-Plas, F., Casieri, L. and Wipf, D. (2012) Sugar transporters in plants and in their interactions with fungi. *Trends Plant Sci.* **17**, 413–422.

Doyle, J.J. and Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus (Madison).* **12**, 13–15.

Ecker, J.R. and Davis, R.W. (1987) Plant defense genes are regulated by ethylene. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 5202–5206.

Esh, A. (2014) "The Activity of Pathogenesis Related Proteins in Smut Resistant and Susceptible Sugarcane (GT54-9) Mutants Induced by Gamma Radiation." *Adv. Plants Agric. Res.* **1**, 1–12.

Fabro, G., Kovács, I., Pavet, V., Szabados, L. and Alvarez, M.E. (2004) Proline accumulation and AtP5CS2 gene activation are induced by plant-pathogen incompatible interactions in Arabidopsis. *Mol. Plant. Microbe. Interact.* **17**, 343–50.

Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–61.

Fontaniella, B., Márquez, A., Rodríguez, C.W., Piñón, D., Solas, M.T., Vicente, C. and Legaz, M.E. (2002) A role for sugarcane glycoproteins in the resistance of sugarcane to Ustilago scitaminea. *Plant Physiol. Biochem.* **40**, 881–889.

Galili, G., Amir, R. and Fernie, A.R. (2016) The Regulation of Essential Amino Acid Synthesis and Accumulation in Plants. *Annu. Rev. Plant Biol.* **67**, 153–178.

Glassop, D., Roessner, U., Bacic, A. and Bonnett, G.D. (2007) Changes in the Sugarcane Metabolome with Stem Development . Are They Related to Sucrose Accumulation ? *Plant Cell Physiol.* **48**, 573–584.

Gullberg, J., Jonsson, P., Nordström, A., Sjöström, M. and Moritz, T. (2004) Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal. Biochem.* **331**, 283–295.

Guo, H. and Ecker, J.R. (2004) The ethylene signaling pathway: new insights. *Curr. Opin. Plant Biol.* **7**, 40–49.

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L. and Bino, R. (2002) Plant metabolomics: the missing link in functional genomics strategies. *Plant Cell* **14**, 1437–40.

Hayat, S., Hayat, Q., Alyemeni, M.N., Wani, A.S., Pichtel, J. and Ahmad, A. (2012) Role of proline under changing environments: A review. *Plant Signal. Behav.* **7**, 1456–1466.

Heinze, B., Thokoane, L., Williams, N., Barnes, J. and Rutherford, R. (2001) The smut-sugarcane interaction as a model system for the integration of marker discovery and gene isolation. *Proc. South African Sugar Technol. Assoc.* **75**, 104–107.

Hincha, D.K., Zuther, E. and Heyer, A.G. (2003) The preservation of liposomes by raffinose family oligosaccharides during drying is mediated by effects on fusion and lipid phase transitions. *Biochim. Biophys. Acta* **1612**, 172–7.

Hoorn, R.A.L. van der and Kamoun, S. (2008) From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009–17.

Huang, N., Zhang, Y.Y., Xiao, X.H., Huang, L., Wu, Q.B., Que, Y.X. and Xu, L.P. (2015) Identification of smut-responsive genes in sugarcane using cDNA-SRAP. *Genet Mol Res* **14**, 6808–6818.

Iskandar, H., Simpson, R., Casu, R., Bonnett, G., Maclean, D. and Manners, J. (2004) Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. *Plant Mol. Biol. Report.* **22**, 325–337.

Jorge, T.F., Mata, A.T. and António, C. (2016) Mass spectrometry as a quantitative tool in plant metabolomics. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374**, 1–26.

Kim, M.S., Cho, S.M., Kang, E.Y., et al. (2008) Galactinol Is a Signaling Component of the Induced Systemic Resistance Caused by Pseudomonas chlororaphis O6 Root Colonization. *Mol. Plant-Microbe Interact.* **21**, 1643–1653.

LaO, M., Arencibia, A., Carmona, E., Acevedo, R., Rodríguez, E., León, O. and Santana, I. (2008) Differential expression analysis by cDNA-AFLP of Saccharum spp. after inoculation with the host pathogen Sporisorium scitamineum. *Plant Cell Rep.* **27**, 1103–1111.

Leme, G.M., Coutinho, I.D., Creste, S., Hojo, O., Carneiro, R.L., Bolzani, V. da S. and Cavalheiro, A.J. (2014) HPLC-DAD method for metabolic fingerprinting of the phenotyping of sugarcane genotypes. *Anal. Methods* **6**, 7781–7788.

Less, H., Angelovici, R., Tzin, V. and Galili, G. (2011) Coordinated Gene Networks Regulating Arabidopsis Plant Metabolism in Response to Various Stresses and Nutritional Cues. *Plant Cell* **23**, 1264–1271.

López-Gresa, M.P., Maltese, F., Bellés, J.M., Conejero, V., Kim, H.K., Choi, Y.H. and Verpoorteb, R. (2010) Metabolic response of tomato leaves upon different plantâ€“pathogen interactions. *Phytochem. Anal.* **21**, 89–94.

Maeda, H. and Dudareva, N. (2012) The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annu. Rev. Plant Biol.* **63**, 73–105.

Mahmud, I., Shrestha, B., Boroujerdi, A. and Chowdhury, K. (2015) NMR-based metabolomics profile comparisons to distinguish between embryogenic and non-embryogenic callus tissue of sugarcane at the biochemical level. *Vitr. Cell. Dev. Biol. - Plant* **51**, 340–349.

McCormick, A.J., Watt, D.A. and Cramer, M.D. (2009) Supply and demand: sink regulation of sugar accumulation in sugarcane. *J. Exp. Bot.* **60**, 357–364.

Millanes, A.M., Fontaniella, B., Legaz, M.E. and Vicente, C. (2005) Glycoproteins from sugarcane plants regulate cell polarity of Ustilago scitaminea teliospores. *J. Plant Physiol.* **162**, 253–265.

Moghaddam, M.R.B. and Ende, W. Van Den (2012) Sugars and plant innate immunity. *J. Exp. Bot.* **63**, 3989–3998.

Papini-Terzi, F.S. (2005) Transcription Profiling of Signal Transduction-Related Genes in Sugarcane Tissues. *DNA Res.* **12**, 27–38.

Peters, L.P. (2016) *A more detailed view of reactive oxygen species metabolism in the sugarcane and Sporisorium scitamineum interaction*. University of São Paulo.

Pfaffl, M.W., Horgan, G.W. and Dempfle, L. (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36–e36.

Piñon, D., Armas, R. de, Vicente, C. and Legaz, M.E. (1999) Role of polyamines in the infection of sugarcane buds by Ustilago scitaminea spores. *Plant Physiol. Biochem.* **37**, 57–64.

Qamar, A., Mysore, K.S. and Senthil-Kumar, M. (2015) Role of proline and pyrroline-5-carboxylate metabolism in plant defense against invading pathogens. *Front. Plant Sci.* **6**, 503.

Que, Y., Xu, L., Lin, J., Ruan, M., Zhang, M. and Chen, R. (2011) Differential protein expression in sugarcane during sugarcane- sporisorium scitamineum interaction revealed by 2-DE and MALDI-TOF-TOF/MS. *Comp. Funct. Genomics* **2011**.

Rae, A.L., Grof, C.P.L., Casu, R.E. and Bonnett, G.D. (2005) Sucrose accumulation in the sugarcane stem: pathways and control points for transport and compartmentation. *F. Crop. Res.* **92**, 159–168.

Ramakers, C., Ruijter, J.M., Deprez, R.H.L. and Moorman, A.F.. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66.

Robinson-Beers, K. and Evert, R. (1991) Ultrastructure of and plasmodesmatal frequency in mature leaves of sugarcane. *Planta* **184**, 291–306.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. and Fernie, A. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11–29.

Rojas, C.M., Senthil-Kumar, M., Tzin, V. and Mysore, K.S. (2014) Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense. *Front. Plant Sci.* **5**, 1–12.

Roje, S. (2006) S-Adenosyl-l-methionine: Beyond the universal methyl group donor. *Phytochemistry* **67**, 1686–1698.

Santiago, R., Alarcón, B., Armas, R. de, Vicente, C. and Legaz, M.E. (2012) Changes in cinnamyl alcohol dehydrogenase activities from sugarcane cultivars inoculated with Sporisorium scitamineum sporidia. *Physiol. Plant.* **145**, 245–259.

Santiago, R., Armas, R. de, Blanch, M., Vicente, C. and Legaz, M.-E. (2010) In vitro effects of caffeic acid upon growth of the fungi Sporisorium scitamineum. *J. Plant Interact.* **5**, 233–240.

Santiago, R., Armas, R. de, Fontaniella, B., Vicente, C. and Legaz, M.-E. (2009) Changes in soluble and cell wall-bound hydroxycinnamic and hydroxybenzoic acids in sugarcane cultivars inoculated with Sporisorium scitamineum sporidia. *Eur. J. Plant Pathol.* **124**, 439–450.

Schaker, P.D.C., Palhares, A.C., Taniguti, L.M., et al. (2016) RNAseq Transcriptional Profiling following Whip Development in Sugarcane Smut Disease Stajich, J.E., ed. *PLoS One* **11**, e0162237.

Sengupta, S., Mukherjee, S., Basak, P. and Majumder, A.L. (2015) Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Front. Plant Sci.* **6**, 1–11.

Song, Z., Cox, R.J., Lazarus, C.M. and Simpson, T.J. (2004) Fusarin C Biosynthesis in Fusarium moniliforme and Fusarium venenatum. *ChemBioChem* **5**, 1196–1203.

Su, Y., Guo, J., Ling, H., et al. (2014) Isolation of a Novel Peroxisomal Catalase Gene from Sugarcane, Which Is Responsive to Biotic and Abiotic Stresses Liu, J.-H., ed. *PLoS One* **9**, e84426.

Su, Y., Xu, L., Wang, Z., Peng, Q., Yang, Y., Chen, Y. and Que, Y. (2016) Comparative proteomics reveals that central metabolism changes are associated with resistance against Sporisorium scitamineum in sugarcane. *BMC Genomics* **17**, 800.

Su, Y., Xu, L., Xue, B., Wu, Q., Guo, J., Wu, L. and Que, Y. (2013) Molecular cloning and characterization of two pathogenesis-related β-1,3-glucanase genes ScGluA1 and ScGluD1 from sugarcane infected by Sporisorium scitamineum. *Plant Cell Rep.* **32**, 1503–1519.

Sundar, A.R., Barnabas, E.L., Malathi, P. and Viswanathan, R. (2012) A Mini-Review on Smut Disease of Sugarcane Caused by Sporisorium scitamineum. In Botany. (Mworia, J.K., ed), pp. 107–128. InTech.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–9.

Taniguti, L.M., Schaker, P.D.C., Benevenuto, J., et al. (2015) Complete Genome Sequence of Sporisorium scitamineum and Biotrophic Interaction Transcriptome with Sugarcane. *PLoS One* **10**, e0129318.

Verslues, P.E. and Sharma, S. (2010) Proline metabolism and its implications for plant-environment interaction. *Arab. B.* **8**, e0140.

Vos, R.C. De, Moco, S., Lommen, A., Keurentjes, J.J., Bino, R.J. and Hall, R.D. (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* **2**, 778–791.

Wang, J., Nayak, S., Koch, K. and Ming, R. (2013) Carbon partitioning in sugarcane (Saccharum species). *Front. Plant Sci.* **4**, 1–6.

Weckwerth, W., Loureiro, M.E., Wenzel, K. and Fiehn, O. (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl. Acad. Sci.* **101**, 7809–7814.

Xia, J., Sinelnikov, I. V., Han, B. and Wishart, D.S. (2015) MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res.* **43**, W251–W257.

You-Xiong, Q., Jian-Wei, L., Xian-Xian, S., Li-Ping, X. and Ru-Kai, C. (2011) Differential Gene Expression in Sugarcane in Response to Challenge by Fungal Pathogen Ustilago scitaminea Revealed by cDNA-AFLP. *J. Biomed. Biotechnol.* **2011**, 1–10.

6. FINAL CONSIDERATIONS

In the present study, it was determined global responses concerning sugarcane smut disease in both plant and pathogen sides. Transcriptome data analysis revealed genes from *S. scitamineum* that are good candidates acting as effector in early moments of the interaction or related to sporulation. Also, the repertoire of genes activated during growth *in planta* suggests several mechanisms that confer advantages to the pathogen, such as cell-wall degrading enzymes, nutrient transporters and detoxification enzymes. Genes encoding to putative effectors target several plant compartments, and their characterization in model systems will allow to determine more precisely mechanisms involved in pathogen recognition or signalization to susceptibility.

Sugarcane responses at transcriptional levels suggest a premature transcriptional reprogramming of the shoot meristem functions continuing until the emergence of the whip. The guidance of this altered pattern is potentially related primarily to auxin mobilization in addition to the involvement of other hormonal imbalances. Several MADS-type transcription factors were up-regulated, indicating that the development of the whip can use a route similar to flowering. Genes encoding RGAs were differentially expressed and may be related to pathogen effector's recognition.

Metabolomics data from sugarcane-smut interaction supported many hypotheses built on transcriptome data, such as the increased energy-related pathways, and the accumulation of starch to feed whip development. Additionally, disease progression was characterized by a shift in the primary metabolism between 65 and 100 DAI, especially of those metabolites related to cell wall biosynthesis, suggesting loosening of the cell wall to allow whip growth. However, increased levels of tyrosine in infected plants may be related of differential PTAL gene expression possibly leading to the synthesis of lignin to feed whip development. Metabolomics also allowed the identification of fungal metabolites, opening a new opportunity in the study of sugarcane smut pathogen, and provided some biomarkers worth to be exploited.

APPENDIX A.

Protein sequences udes to build PAL/PTAL phylogenetic tree.

>comp192201_c0_seq1_Sugarcane_unigenes
EENLKSAVKNSVMAVARKVLTTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEAEKDASASVFSKINRFEETLREVLPREME
AARVAFETGTAPIANRIKESRSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC

>SHCRBa_019_F13_F_30_1_bac_Cana
MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEAEKDASASVFSKINRFEETLREVLPREMEAARVAFETGTAPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC

>comp94629_c0_seq1_Sugarcane_unigenes
MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGF

>750133730_Saccharum_officinarum_barnabas
XXXKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFX

>comp202731_c0_seq1_Sugarcane_unigenes
MECENGHVAAASGNGVCLATPRAADPLNWGKAAEDLTGSHLEAVKRMVEEYRRPLVKIEGGSLTVAQVAAVAAAGEARVELDESARGRVKASSDWVMSSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAGLI
TGRQNSVAVAPDGRKVDAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLATMAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMTSTFLIALCQAIDLRHIEENVKAAVK
NCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDAERVVETSIFAKVAEFEQQVRAALPKEVEAARAAVESG
NPLVPNRIKECRSYPLYRFVREEVGTEYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKEWNGEPLPLC

>gg_14643_Sugarcane_unigenes
MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLTVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAP

>Q96V77_Ustilago_maydis
MAPTADVLPPVEASTRPGLLVQPSDTKLRKASSFRTEQVVIDGYNLKIQGLVASARYGHVPVLDPSAETRKRIDDSVQSLIAKLDRGESIYGINTGFGGSADSRTAN
TRALQLALLQMQQCGVLPVPSTFPTGEPSSAPFALPLTDTESSLIMPEAWVRGAIVVRLSSLMRGHSGVRWEVLDKMQKLFLQNNVTPVVPVRSSISASGDLSPLSY
VAGALAGQRGIYCFVTDGRGQRVKVTADEACRMHKITPVQYEPKEALGLLNGTAFSASVAGLATYEAENLASLTQLTTAMAVEALKGTDASFAPFIHEIARPHPGQI
KSAKFIRALLSGSRLAEHLENEKHVLFSEDNGTLRQDRYTLRTASQWVGPGLEDIENAKRSVDIEINSTTDNPMIDPYDGDGRIHHGGNFQAMAMTNAVEKIRLALC
AMGKMTFQQMTELVNPAMNRGLPANLASTPDLSLNFHAKGIDIALASVTSELMFLGNPVSTHVQSAEMANQAINSLALISGRQTLQAIECLSMIQAWSLYLLCQALD
IRALQYKVAEQLPTLILASLHSHFGEWMDETKQQEIAAQVLKSMSKRLDETSSKDLRDRLVETYQDASSVLVRYFSELPSGGGADPLRNIVKWRATGVADTEKIYRQ
VTIEFLDNPYACHASHLLGKTKRAYEFVRKTLGVPMHGKENLNEFKGEFEQWNTTGGYVSVIYASIRDGELYNMLSELERDL

>A0A0F7SDN6_Sporisorium_scitamineum
MAPTADVLPAVETCARPGLLVQLCDTKIRKASSFRTEQVVIDGNNLGIQGLVASARYGHVPVLDASSATRKRIDDSVQSLIAKLDRGESIYGINTGFGGSADSRTAN
TRALQLALLQMQQCGVLPVPSTFPTGEPSSAPFALPLTDTESSLIMPEAWVRGAIVVRLSSLMRGHSGVRWEVLDKMQKLFLQNNVTPVVPVRSSISASGDLSPLSY
VAGALAGQRGIYCFVTGKNGQRVKVTADEACRMHNITPVQYEPKEALGLLNGTAFSASVAGLATYEAEKLAALTQLTTAMAVEALKGTDASFAPFIHEIARPHPGQI
KSAKFIRALLSGSKLAEHLENEKHVLFSEDNGTLRQDRYTLRTASQWVGPGLEDIENAKRSVDIEINSTTDNPMIDPYDGDGRIHHGGNFQAMAMTNAVEKIRLALC
AMGKMTFQQMTELVNPAMNRGLPANLASTPDLSLNFHAKGIDIALASVTSELMFLGNPVSTHVQSAEMANQAINSLALISGRQTLQAVECLSMIQAWSLYLLCQALD
IRALQHKVAEQLPALILASLHAHFGEWMDKDKQLEIAAQVLKSMSKRLDETSSKDLRDRLVETYQDASSVLVKYFSELPSGGGADPLRNIVKWRAAGVADTEKIYRQ
VTIEFLDNPHACHASHLLGKTKRAYEFVRKTLGVPMHGKENLNEFKGEFEQWNTTGGYVSVIYASIRDGELYDMLGELEADL

>R9P5K0_Pseudozyma_hubeiensis
MAPTADVLPAAEASARPGLLVQPSDTKLRKASSFRTEQVVIDGYNLKIQGLVASARYGHVPVLDSSAATRKRIDDSVQSLIAKLDRGESIYGINTGFGGSADSRTAN
TRALQLALLQMQQCGVLPVPSTFPTGEPSSAPFALPLTDTESSLIMPEAWVRGAIVVRLSSLMRGHSGVRWEVLEKMQKLFLENNVTPVVPVRSSISASGDLSPLSY
VAGALAGQRGIYCFVTDQRGQRVRVTADEACRMHNITPVQYEPKEALGLLNGTAFSASVAGLATYEAEKLAALTQLTTAMAVEALKGTDASFAPFIHEIARPHPGQI
KSAKFIRALLSGSKLALHLEDEKHVLFSEDNGTLRQDRYTLRTASQWVGPGLEDIENAKRSVDIEINSTTDNPMIDPYDGDGRIHHGGNFQAMAMTNAVEKIRLALC
AMGKMTFQQMTELVNPAMNRGLPANLASTPDLSLNFHAKGIDIALASVTSELMFLGNPVSTHVQSAEMANQAINSLALISGRQTLQAVECLSMIQAWSLYILCQAYD
IRALQHKVAEQLPALILGSLRVHFGEWMDEAKQQEIAVLVLKSMSKRLDETSSKDLRDRLVETYQDASSVLVKYFSELPSGGGADPLRNIVKWRAAGVLDTEAIYRN
VTVEFLDNPHACHASHLLGKTKRAYEFVRKTLGVPMHGKENLNEFRGEFEQWNTTGGYVSVIYASIRDGELYSMLGELERDL

>E6ZJF9_Sporisorium_reilianum_SRZ2
MAPTADVLPAVEASPRPGLLVQPSDTKIRKASSFRTEQVVIDGNNLKIQGLVASARYGHVPILDASPATRKRIDDSVHSLLAKLDRGESIYGINTGFGGSADSRTAN
TRALQLALLQMQQCGVLPVPSSFPTGEPSSAPFALPLTDTESSLIMPEAWVRGAIVVRLSSLMRGHSGVRWEVLDKMQRLFLENNVTPVVPVRSSISASGDLCPLSY
VAGALAGQRGIYCFVTDNHGQRVKVTADEACRMHNITPVQYEPKEALGLLNGTAFSASVAGLATYEADKLAALTQLTTAMAVEALKGTDASFAPFIHEVARPHPGQI
KSAKFIRALLSGSKLAEHLENEKHVLFSEDNGTLRQDRYTLRTASQWVGPGLEDIENAKRSVDIEINSTTDNPMIDPYDGDGRIHHGGNFQAMAMTNAVEKIRLALC
AMGKMTFQQMTELVNPAMNRGLPANLASTPDLSLNFHAKGIDIALASVTSELMFLGNPVSTHVQSAEMANQAINSLALISGRQTLQAVECLSMIQAWSLYLLCQALD
IRALQHKVAEQLPALLLASLHAHFGEWMDNDTQLAIAAQVLKSMSRRLDETSSKDLRDRLVETYQDASSVLVKYFSELPSGGGADPLRNIVKWRAAGVADTEKIYRD
VTIEFLDNPHACHASHLLGKTRRAYEFVRKTLGVPMHGKENLNEFKGEFEQWNTTGGYVSVIYASIRDGELYNMLSQLEGDL

>I2G3P6_Ustilago_hordei
MAPTAHVLPMPIPTETRPSLAVQPSDTKLRKASSFRTEQVVIDGNNLKIQGLVASARYGHVPVLDSSPAIRKRIDDSVQSLIAKLDRGESIYGINTGFGGSADSRTA
NTRALQLALLQMQQCGVLPVPSTFPTGEPSSAPFALPLTDTETSLVMPEAWVRGAIVVRLSSLMRGHSGVRWQVLEKMQKLFLQNNVTPVVPVRSSISASGDLSPLS
YVAGALAGQKGIYCWITDAKSGQRVKVTADEACRMYAIEPVQYEPKEALGLLNGTAFSASVAALATYEAEKLANLTQLTTAMAVEALKGTDASFAPFIHQVARPHPG
QIKSARFIRALLSGSQLAEHLENEKHVLFSEDNGTLRQDRYTLRTASQWVGPGLEDIENAKRSVDIEINSTTDNPMIDPYDGDGRIHHGGNFQAMAMTNAVEKIRLA
LCAMGKMTFQQMTELVNPAMNRGLPANLASTPDLSLNFHAKGIDIALASVTSELMFLGNPVSTHVQSAEMANQAINSLALISGRMTLQAVECLSMIQAWSLYLLCQA
LDIRALQHKVAEQLPALILASLNSHFGEWMDEAKQAEIAKLVLKQMSKRLDETSSKDLRDRLVETYQDASSVLVKYFSELPSGGGADPLRNIVKWRAAGVADTEKIY
RDVTVEFLDNPYACHASHLLGKTKRAYEFVRKTLGVPMHGKENLNEFKGEFSQWNTTGGYVSVIYASIRDGELYSMLSELEKDLQL

>B2J528_Nostoc_punctiforme
MNITSLQQNITRSWQIPFTNSSDSIVTVGDRNLTIDEVVNVARHGTQVRLTDNADVIRGVQASCDYINNAVETAQPIYGVTSGFGGMADVVISREQAAELQTNLIWF
LKSGAGNKLSLADVRAAMLLRANSHLYGASGIRLELIQRIETFLNAGVTPHVYEFGSIGASGDLVPLSYITGALIGLDPSFTVDFDGKEMDAVTALSRLGLPKLQLQ
PKEGLAMMNGTSVMTGIAANCVYDAKVLLALTMGVHALAIQGLYGTNQSFHPFIHQCKPHPGQLWTADQMFSLLKDSSLVREELDGKHEYRGKDLIQDRYSLRCLAQ
FIGPIVDGVSEITKQIEVEMNSVTDNPLIDVENQVSYHGGNFLGQYVGVTMDRLRYYIGLLAKHIDVQIALLVSPEFSNGLPPSLVGNSDRKVNMGLKGLQISGNSI

MPLLSFYGNSLADRFPTHAEQFNQNINSQGYISANLTRRSVDIFQNYMAIALMFGVQAVDLRTYKMKGHYDARTCLSPNTVQLYTAVCEVVGKPLTSVRPYIWNDNE
QCLDEHIARISADIAGGGLIVQAVEHIFSSLKST

>P11544_Rhodosporidium_toruloides

MAPSLDSISHSFANGVASAKQAVNGASTNLAVAGSHLPTTQVTQVDIVEKMLAAPTDSTLELDGYSLNLGDVVSAARKGRPVRVKDSDEIRSKIDKSVEFLRSQLSM
SVYGVTTGFGGSADTRTEDAISLQKALLEHQLCGVLPSSFDSFRLGRGLENSLPLEVVRGAMTIRVNSLTRGHSAVRLVVLEALTNFLNHGITPIVPLRGTISASGD
LSPLSYIAAAISGHPDSKVHVVHEGKEKILYAREAMALFNLEPVVLGPKEGLGLVNGTAVSASMATLALHDAHMLSLLSQSLTAMTVEAMVGHAGSFHPFLHDVTRP
HPTQIEVAGNIRKLLEGSRFAVHHEEEVKVKDDEGILRQDRYPLRTSPQWLGPLVSDLIHAHAVLTIEAGQSTTDNPLIDVENKTSHHGGNFQAAAVANTMEKTRLG
LAQIGKLNFTQLTEMLNAGMNRGLPSCLAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLLATHLYCVLQAI
DLRAIEFEFKKQFGPAIVSLIDQHFGSAMTGSNLRDELVEKVNKTLAKRLEQTNSYDLVPRWHDAFSFAAGTVVEVLSSTSLSLAAVNAWKVAAAESAISLTRQVRE
TFWSAASTSSPALSYLSPRTQILYAFVREELGVKARRGDVFLGKQEVTIGSNVSKIYEAIKSGRINNVLLKMLA

>Q87V42_Pseudomonas_syringae

MTNQVPDPIMFGERALCIEDVLALANRQAPSALQGDDEFRARIARGAQFLDSLLSKEGVIYGVTTGYGDSCVVAVPLEHVEALPRYLYTFHGCGLGKLLDAQATRAV
LAARLQSLCQGVSGVRVELLERLQAFIDQDVLPLIPEEGSVGASGDLTPLSYVAATLSGEREVMFRGERRLASDVHRELGWTPLVLRPKEALALMNGTAVMTGIACL
AFARADYLLQLATRITAMNVVALQGNPEHFDERLFAAKPHPGQMQVAAWLRQDLAIDAPTAPLHRLQDRYSLRCAPHVLGVLADSLNWLRSFIEIELNSANDNPIID
AEAERVLHGGHFYGGHIAFAMDSLKTLVANVADLLDRQLALLVDERYNHGLPSNLSGASAERAMLNHGFKAVQIGTSAWTAEALKNTMPASVFSRSTECHNQDKVSM
GTIAARDAIRVLELTEQVAAATLIAANQGVWLRSKGADARPLPPALASMHAELGEDFAPVIEDRALESELRLCLKHIANRRWRLHAQ

>A0A099SNC3_Pseudomonas_syringae

MTNQVPDPIMFGERALCIEDVLALANRQAPSALQGDDEFRARIARGAQFLDSLLSKEGVIYGVTTGYGDSCVVAVPLEHVEALPRYLYTFHGCGLGKLLDAQATRAV
LAARLQSLCQGVSGVRVELLERLQAFIDQDVLPLIPEEGSVGASGDLTPLSYVAATLSGEREVMFRGERRLASDVHRELGWTPLVLRPKEALALMNGTAVMTGIACL
AFARADYLLQLATRITAMNVVALQGNPEHFDERLFAAKPHPGQMQVAAWLRQDLAIDAPTAPLHRLQDRYSLRCAPHVLGVLADSLNWLRSFIEIELNSANDNPIID
AEAERVLHGGHFYGGHIAFAMDSLKTLVANVADLLDRQLALLVDERYNHGLPSNLSGASAERAMLNHGFKAVQIGTSAWTAEALKNTMPASVFSRSTECHNQDKVSM
GTIAARDAIRVLELTEQVAAATLIAANQGVWLRSKGADARPLPPALASMHAELGEDFAPVIEDRALESELRLCLKHIANRRWRLHAQ

>A0A0W0KF34_Pseudomonas_syringae

MTNQVPDPIMFGERALCIEDVLALANRQAPSALQGDDEFRARIARGAQFLDSLLSKEGVIYGVTTGYGDSCVVAVPLEHVEALPRYLYTFHGCGLGKLLDAQATRAV
LAARLQSLCQGVSGVRVELLERLQAFIDQDVLPLIPEEGSVGASGDLTPLSYVAATLSGEREVMFRGERRLASDVHRELGWTPLVLRPKEALALMNGTAVMTGIACL
AFARADYLLQLATRITAMNVVALQGNPEHFDERLFAAKPHPGQMQVAAWLRQDLAIDAPTAPLHRLQDRYSLRCAPHVLGVLADSLNWLRSFIEIELNSANDNPIID
AEAERVLHGGHFYGGHIAFAMDSLKTLVANVADLLDRQLALLVDERYNHGLPSNLSGASAERAMLNHGFKAVQIGTSAWTAEALKNTMPASVFSRSTECHNQDKVSM
GTIAARDAIRVLELTEQVAAATLIAANQGVWLRSKGADARPLPPALASMHAELGEDFAPVIEDRALESELRLCLKHIANRRWRLHAQ

>Q8GMG0_Streptomyces_globisporus

MALTQVETEIVPVSVDGETLTVEAVRRVAEERATVDVPAESIAKAQKSREIFEGIAEQNIPIYGVTTGYGEMIYMQVDKSKEVELQTNLVRSHSAGVGPLFAEDEAR
AIVAARLNTLAKGHSAVRPIILERLAQYLNEGITPAIPEIGSLGASGDLAPLSHVASTLIGEGYVLRDGRPVETAQVLAERGIEPLELRFKEGLALINGTSGMTGLG
SLVVGRALEQAQQAEIVTALLIEAVRGSTSPFLAEGHDIARPHEGQIDTAANMRALMRGSGLTVEHADLRRELQKDKEAGKDVQRSEIYLQKAYSLRAIPQVVGAVR
DTLYHARHKLRIELNSANDNPLFFEGKEIFHGANFHGQPIAFAMDFVTIALTQLGVLAERQINRVLNRHLSYGLPEFLVSGDPGLHSGFAGAQYPATALVAENRTIG
PASTQSVPSNGDNQDVVSMGLISARNARRVLSNNNKILAVEYLAAAQAVDISGRFDGLSPAAKATYEAVRRLVPTLGVDRYMADDIELVADALSRGEFLRAIARETD
IQLR

>P21310_Pseudomonas_putida

MTELTLKPGTLTLAQLRAIHAAPVRLQLDASAAPAIDASVACVEQIIAEDRTAYGINTGFGLLASTRIASHDLENLQRSLVLSHAAGIGAPLDDDLVRLIMVLKINS
LSRGFSGIRRKVIDALIALVNAEVYPHIPLKGSVGASGDLAPLAHMSLVLLGEGKARYKGQWLSATEALAVAGLEPLTLAAKEGLALLNGTQASTAYALRGLFYAED
LYAAAIACGGLSVEAVLGSRSPFDARIHEARGQRGQIDTAACFRDLLGDSSEVSLSHKNCDKVQDPYSLRCQPQVMGACLTQLRQAAEVLGIEANAVSDNPLVFAAE
GDVISGGNFHAEPVAMAADNLALAIAEIGSLSERRISLMMDKHMSQLPPFLVENGGVNSGFMIAQVTAAALASENKALSHPHSVDSLPTSANQEDHVSMAPAAGKRL
WEMAENTRGVLAIEWLGACQGLDLRKGLKTSAKLEKARQALRSEVAHYDRDRFFAPDIEKAVELLAKGSLTGLLPAGVLPSL

>Q3IWB0_Rhodobacter_sphaeroides

MLAMSPPKPAVELDRHIDLDQAHAVASGGARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGPVLDWTTARAMV
LARLVSIAQGASGASEGTIARLIDLLNSELAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTRLDGAEGLRRGRLQPLDLSHRDALALVNGTSAMTGIALV
NAHACRHLGNWAVALTALLAECLRGRTEAWAAALSDLRPHPGQKDAAARLRARVDGSARVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDTLAWHDR
VLTIELNAVTDNPVFPPDGSVPALHGGNFMGQHVALTSDALATAVTVLAGLAERQIARLTDERLNRGLPPFLHRGPAGLNSGFMGAQVTATALLAEMRATGPASIHS
ISTNAANQDVVSLGTIAARLCREKIDRWAEILAILALCLAQAAELRCGSGLDGVSPAGKKLVQALREQFPPLETDRPLGQEIAALATHLLQQSPV

>Q9SS45_Arabidopsis_thaliana

MELCNQNNHITAVSGDPLNWNATAEALKGSHLDEVKRMVKEYRKEAVKLGGETLTIGQVAAVARGGGGSTVELAEEARAGVKASSEWVMESMNRGTDSYGVTTGFGA
TSHRRTKQGGALQNELIRFLNAGIFGPGAGDTSHTLPKPTTRAAMLVRVNTLLQGYSGIRFEILEAITKLLNHEITPCLPLRGTITASGDLVPLSYIAGLLTGRPNS
KAVGPSGETLTASEAFKLAGVSSFFELQPKEGLALVNGTAVGSGLASTVLFDANILAVLSEVMSAMFAEVMQGKPEFTDHLTHKLKHHPGQIEAAAIMEHILDGSSY
VKEAQLLHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRAATKMIEREINSVNDNPLIDVSRNKALHGGNFQGTPIGVAMDNSRLAIASIGKLMFAQFSELVNDFYNN
GLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSTTYLVALCQAVDLRHLEENLKKAVKSAVSQVA
KRVLTVGANGELHPSRFTERDVLQVVDREYVFSYADDPCSLTYPLMQKLRHILVDHALADPEREANSATSVFHKIGAFEAELKLLLPKEVERVRVEYEEGTSAIANR
IKECRSYPLYRFVRDELNTELLTGENVRSPGEEFDKVFLAISDGKLIDPLLECLKEWNGAPVSIC

>P45729_Petroselinum_crispum

MAYVNGTTNGHANGNGLDLCMKKEDPLNWGVAAEALTGSHLDEVKRMVAEYRKPVVKLEGETLTISQVAAISARDDSGVKVELSEEARAGVKASSDWVMDSMNKGTD
SYGVTTGFGATSHRRTKQGGALQKELIRFLNAGIFGSGAEAGNNTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNHNITPCLPLRGTITASGDLVPLSYI
AGLLTGRPNSKAVGPTGVTLSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEVMSAIFAEVMQGKPEFTDHLTHKLKHHPGQIEAAA
IMEHILDGSAYVKAAQKLHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGSPIGVSMDNTRLAIAAIGKLMFAQ
FSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAIDLRHLEENLK
STVKNTVSQVAKRVLTMGVNGELHPSRFCEKDLLRVVDREYIFAYIDDPCSATYPLMQKLRETLVEHALNNGDKERNLSTSIFQKIAAFEDELKALLPKEVETARAA
LESGNPAIPNRIKECRSYPLYKFVREELGTEYLTGEKVRSPGEEFEKVFTAMSKGEIIDPLLECLESWNGAPLPIC

>P24481_Petroselinum_crispum

MENGNGATTNGHVNGNGMDFCMKTEDPLYWGIAAEAMTGSHLDEVKKMVAEYRKPVVKLGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDWVMDSMNKGT
DSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGIFGNGSDNTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPLSYIAG
LLTGRPNSKAVGPTGVILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEVMSAIFAEVMQGKPEFTDHLTHKLKHHPGQIEAAAIM
EHILDGSAYVKAAQKLHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFS
ELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAIDLRHLEENLKST
VKNTVSSVAKRVLTMGVNGELHPSRFCEKDLLRVVDREYIFAYIDDPCSATYPLMQKLRQTLVEHALKNGDNERNLSTSIFQKIATFEDELKALLPKEVESARAALE
SGNPAIPNRIEECRSYPLYKFVRKELGTEYLTGEKVTSPGEEFEKVFIAMSKGEIIDPLLECLESWNGAPLPIC

>P45728_Petroselinum_crispum

MENGNGAITNGHVNGNGMDFCMKTEDPLYWGIAAEAMTGSHLDEVKKMVAEYRKPVVKLGGETLTISQVAAISARDGSGVTVELSEAARAGVKASSDWVMDSMNKGT
DSYGVTTGFGATSHRRTKQGGALQKELIRFLNAGIFGNGSDNTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKFLNQNITPCLPLRGTITASGDLVPLSYIAG
LLTGRPNSKAVGPTGVILSPEEAFKLAGVEGGFFELQPKEGLALVNGTAVGSGMASMVLFEANILAVLAEVMSAIFAEVMQGKPEFTDHLTHKLKHHPGQIEAAAIM
EHILDGSAYVKAAQKLHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRSSTKMIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGMSMDNTRLAIAAIGKLMFAQFS
ELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTSEAVEILKLMSTTFLVGLCQAIDLRHLEENLKST
VKNTVSSVAKRVLTMGVNGELHPSRFCEKDLLRFVDREYIFAYIDDPCSATYPLMQKLRQTLVEHALKNGDNERNMNTSIFQKIATFEDELKALLPKEVESARAALE
SGNPAIPNRIEECRSYPLYKFVRKELGIEYLTGEKVTSPGEEFDKVFIAMSKGEIIDPLLECLESWNGAPLPIC

>P35510_Arabidopsis_thaliana

MEINGAHKSNGGGVDAMLCGGDIKTKNMVINAEDPLNWGAAAEQMKGSHLDEVKRMVAEFRKPVVNLGGETLTIGQVAAISTIGNSVKVELSETARAGVNASSDWVM
ESMNKGTDSYGVTTGFGATSHRRTKNGVALQKELIRFLNAGIFGSTKETSHTLPHSATRAAMLVRINTLLQGFSGIRFEILEAITSFLNNNITPSLPLRGTITASGD

LVPLSYIAGLLTGRPNSKATGPNGEALTAEEAFKLAGISSGFFDLQPKEGLALVNGTAVGSGMASMVLFETNVLSVLAEILSAVFAEVMSGKPEFTDHLTHRLKHHP
GQIEAAAIMEHILDGSSYMKLAQKLHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRYATKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAI
GKLMFAQFSELVNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLVAICQAVDLR
HLEENLRQTVKNTVSQVAKKVLTTGVNGELHPSRFCEKDLLKVVDREQVYTYADDPCSATYPLIQKLRQVIVDHALINGESEKNAVTSIFHKIGAFEEELKAVLPKE
VEAARAAYDNGTSAIPNRIKECRSYPLYRFVREELGTELLTGEKVTSPGEEFDKVFTAICEGKIIDPMMECLNEWNGAPIPIC

>P45724_Arabidopsis_thaliana

MDQIEAMLCGGGEKTKVAVTTKTLADPLNWGLAADQMKGSHLDEVKKMVEEYRRPVVNLGGETLTIGQVAAISTVGGSVKVELAETSRAGVKASSDWVMESMNKGTD
SYGVTTGFGATSHRRTKNGTALQTELIRFLNAGIFGNTKETCHTLPQSATRAAMLVRVNTLLQGYSGIRFEILEAITSLLNHNISPSLPLRGTITASGDLVPLSYIA
GLLTGRPNSKATGPDGESLTAKEAFEKAGISTGFFDLQPKEGLALVNGTAVGSGMASMVLFEANVQAVLAEVLSAIFAEVMSGKPEFTDHLTHRLKHHPGQIEAAAI
MEHILDGSSYMKLAQKVHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRQATKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQF
SELVNDFYNNGLPSNLTASSNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTSEAVDILKLMSTTFLVGICQAVDLRHLEENLRQ
TVKNTVSQVAKKVLTTGINGELHPSRFCEKDLLKVVDREQVFTYVDDPCSATYPLMQRLRQVIVDHALSNGETEKNAVTSIFQKIGAFEEELKAVLPKEVEAARAAY
GNGTAPIPNRIKECRSYPLYRFVREELGTKLLTGEKVVSPGEEFDKVFTAMCEGKLIDPLMDCLKEWNGAPIPIC

>B2Z6R0_Populus_trichocarpa

METVTKNGYQNGSLESLCVNQLDPLSWGVAAEAMKGSHLDEVKRMVADYRKPVVKLGGETLTIAQVASIAGHDTGDVKVELSESARPGVKASSDWVMDSMDKGTDSY
GVTTGFGATSHRRTKQGGALQKELIRFLNAGIFGNGTETCHTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITRLLNNNITPCLPLRGTITASGDLVPLSYIAGL
LTGRPNSKATGPTGEVLDAAEAFKAAGIESGFFELQPKEGLALVNGTAVGSGLASMVLFETNVLAVLSELLSAIFAEVMNGKPEFTDHLTHKLKHHPGQIEAAAIME
HILDGSAYMKAAKKLHETDPLQKPKQDRYALRTSPQWLGPQIEVIRFSTKSIEREINSVNDNPLIDVSRNKAIHGGNFQGTPIGVSMDNVRLAIASIGKLLFAQFSE
LVNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCSELQYLANPVTTHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMSTTFLVALCQAIDLRHLEENLKSAV
KNTVSQVSKRVLTTGANGELHPSRFCEKELLKVVDREYVFAYVDDPCSATYPLMQKLRQVFVDHALENGENEKNFSTSVFQKIEAFEEELKALLPKEVESARAAYDS
GNSAIDNKIKECRSYPLYKFVREELGTVLLTGEKVQSPGEEFDKVFTAMCQGKIIDPMLECLGEWNGSPLPIC

>B2Z6R1_Populus_trichocarpa

METITKNGYQNGSSESLCTQRDPLSWGVAAEAMKGSHLDEVKRMVAEYRKPVVNLAGQTLTIAQVASIAGHDASNVKVELSESARPRVKASSDWVMDSMDKGTDSYG
VTTGFGATSHRRTKQGGALQKELIRFLNAGIFGNGTETCHTLPHSATRAAMLVRINTLLQGYSGIRFEILEAITKLLNNNITPCLPLRGTITASGDLVPLSYIAGLL
TGRPNSKATGPNGEVLDAVEAFKAAGIDSGFFELQPKEGLALVNGTAVGSGLASMVLFETNVLAVLSELISAIFAEVMNGKPEFTDHLTHKLKHHPGQIEAAAIMEH
ILDGSAYMKAAKKLHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRFSTKSIEREINSVNDNPLIDVSRNKALHGGNFQGTPIGVSMDNVRLAIASIGKLLFAQFSEL
VNDFYNNGLPSNLTASRNPSLDYGFKGAEIAMASYCSELQYLANPVTSHVQSAEQHNQDVNSLGLISSRKTAESVDILKLMSTTFLVALCQAIDLRHLEENLRSAVK
NTVSHVSKRVLTTGANGELHPSRFCEKELLKVVDREDVFAYADDPCSATYPLMQKLRQVLVDHALANGENEKNTSTSVFQKITAFEEELKALLPKEVESARAAYDSG
NSAIENKIKECRSYPLYKFVREELGTGLLTGEKVRSPGEEFDKVFTAMCQGKIIDPMLECLGEWNGAPLPIC

>A3AVL7_Oryza_sativa_japonica

MASQTADAHGVRRERPAVVGQGGAGDDGEPPGRGEAHGGAVTGGRGEDRGVQPPRRPGGRRLRGQGRLRRGGRAGRGGPPPRQGQQRVGSSTASPTAATSTASPPAS
AGTSNRRTKDGQALQVELLRHLNAGIFGNGSDGNSLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLINTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNA
QAVTVDGKKVDAAEAFKIAGIQGGFFRLEPKEGLAIVNGTSVGSALAAMVLYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILAGSA
FMPHAQKVNEVDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNTRLAIANIGKLMFAQFSELVNEFYN
NGLTSNLAGSRNPSLDYGFKGTEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYLVALCQAVDLRHLEENLKSAVKNCVTTV
AKKVLTTGPAGGLHSARFSEKALLTAIDREAVYSYADDPCSANYPLMTKIRAVLVEHALANGPAEKDDGSSVFSKITAFEEELREALPREMEAARVAFETGTAPITN
RIKESRSFPLYRFVREELGCVYLTGEKLKSPGEECNKVFLAISERKLIDPMLECLKEWNGEPLPIC

>Q84VE0_Oryza_sativa_japonica

MVAQSREAVVKIEGSSLRVGQVAAVSAAKDASGVVVELDEEARPRVKASSEWILNCIAHGGDIYGVTTGFGGTSHRRTKDGQALQVELLRHLNAGIFGNGSDGNSLP
SEVSRAAMLVRINTLLQGYSGIRFEILEAITKLINTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTVDGKKVDAAEAFKIAGIQGGFFRLEPKEGLAIV
NGTSVGSALAAMVLYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILAGSAFMPHAQKVNEVDPLLKPKQDRYALRTSPQWLGPQIEV
IRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNTRLAIANIGKLMFAQFSELVNEFYNNGLTSNLAGSRNPSLDYGFKGTEIAMASYCSELQFLA
NPVTNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYLVALCQAVDLRHLEENLKSAVKNCVTTVAKKVLTTGPAGGLHSARFSEKALLTAIDREAVYSYAD
DPCSANYPLMTKIRAVLVEHALANGPAEKDDGSSVFSKITAFEEELREALPREMEAARVAFETGTAPITNRIKESRSFPLYRFVREELGCVYLTGEKLKSPGEECNK
VFLAISERKLIDPMLECLKEWNGEPLPIC

>K3YDF1_Setaria_italica

MACSTAIVTSDPLNWGKAAAELTGSHLDEVRRMVAQSREPVVRVDGSRLHVGQVAAVAAAKDASGVAVELDEEARLRVKASSEWVLSCIENGGDIYGVTTGFGGNSH
RRTKDGHALQVELLRYLNAGIFGTGSDGHTLPSQVSRAAMLVRINALMQGYSGIRFEILEAIAKLINTGVSPCLPLRGSITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAMVNGTSVGSALAAMVCFDANVLAVLAVVLSAVFCEVMNGKPEYADHLTHKLKHHPGSIEAAAIMEHILDGSSLMKHA
KEVNAMDPLLKPKQDQYALRTSPQWLGPQIEVIRAATKAIEREINSVSDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAVASIGRLMFAQFTELVIDFYNNGLPS
NLAGSRNLSLDFGLKGAEIAMASYCSELQYLANPVTNHVQSAEQHTQDVNSLGLISARKTAEAVEILKLMSSTFMIALCQAVDLRHLEENLKSAVKNCVKTVALKVL
TTSPDGEHCSARFSEKALLAAIDRKAVYSYYDDPCSASSSLMMTIRAVLVDHALANGEAENEARAPIFSKITKFEEELREALPREMEKTRVAFETGTAPIGNRIKES
RSYPLYRFIREDLGAVYLTGEKLKSAGEECNKVFLALSEGKLIDPMLGCLKEWNGEPLPIC

>K3ZR63_Setaria_italica

MACNTAIVTSDPLNWGKAAAELTGSHLDEVRRMVAQSREPIVRVDGSRLHVGKVAAVAAAKDASGVAVELDEEARLRVRSSSEWVLSCIENGGDIYGVTTGFGGNSH
RRTKDGHALQVELLRYLNAGIFGTGSDGHTLPSQVSRAAMLVRINALMQGYSGIRFEILEAIAKLINTGVSPCLPLRGSITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIKGGFFKLNPKEGLAMVNGTSVGSALAAMVCFDANVLAVLAVVLSAVFCEVMNGKPEYADHLTHKLKHHPGSIEAAAIMEHILDGSSLMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKAIEREINSVSDNPVIDVNRGKALHGGNFQGTPIGVSMDNARLAVASIGRLMFAQFTELVIDFYNNGLPS
NLAGSRNLSLDFGLKGVEIAMASYCSELQYLANPVTNHVQSAEQHTQDVNSLGLISARKTAEAVEILKLMSSTFMIALCQAVDLRHLEENLKSAVKNCVKTVALKVL
TTSPDGEHCSARFSEKTLLAAIDRKAVYSYCDDPCSASSSLMMTIRAVLVDHALANGEAENEARAPIFSKITKFEEELREALPREMEKTRVAFETGTAPIGNRIKES
RSYPLYRFIREDLGAVYLTGEKLKSAGEECNKVFLALSEGKLIDPMLGCLKEWNGEPLPIC

>I1IBR5_Brachypodium_distachyon

MAGNGAISEKDPLNWGAAAAELTGSHLDEVKRMVAQFREPVVKIEGASLRVGQVAAVAQAKDAAGVSVELDEEARPRVKASSEWILSCLAAGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHSLPAEVTRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTA
DGRKVDAAEAFKVAGIEGGFFTLNPKEGLAIVNGTSVGSALAATVLFDCNVLAVLSEVLSAVFCEVMNGKPEFTDHLTHKLKHHPGSIEAAAIMEHILAGSSFMSHA
KKVNEIDPQLKPKQDRYALRTSPQWLGPQIEVIRSATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNTRLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPVTNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENIKASVKNCVTQVSKKVL
TMNPTGDLSSARFSEKSLLTAIDREAVFSYADDACSANYPLMQKLRAVLVDHALTSSGVDNAGESEATVFSKINKFEEELRAALPREIEAARVAFEKGTAPIPNLIK
DSRSFPLYRFVREELGCVYLTGEKLLSPGEECNKVFIGISQGKLIDPMLECLKEWNGEPLPINVV

>D5KS97_Bambusa_oldhamii

MAGNGPIVKDDPLNWGAAAAELTGSHFDEVKRMVAQFREPVIKIEGASLRVGQVAAVAQAKDVSGVAVELDEEARPRVKASSEWILNCLAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGTDGHTLPSEVTRAAMLVRINTLLQGYSGIRFEILEAITKLINTGVTPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVAP
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVLYDCNVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILAGSSFMSHA
KKVNEMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNTRLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENIKSSVKNCVTQVAKKVL
TMNPTGDLSSARFSEKNLLTAIDREAVFTYADDPCSANYPLMQKLRAVLVDHALTSGDAEREPSVFSKITKFEEELRSALPREIEAARVAVADGTAPIANRIKESRS
FPVYRFVREELGCVYLTGEKLKSPGEECNKVFIGISQGKLIDPMLECLKEWNGEPLPIN

>P14717_Oryza_sativa_japonica

MAGNGPINKEDPLNWGAAAAEMAGSHLDEVKRMVAQFREPLVKIQGATLRVGQVAAVAQAKDAAGVAVELDEEARPRVKASSEWILNCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSETVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVTPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAISP
DGRKVDAAEAFKLAGIEGGFFTLNPKEGLAIVNGTSVGSALAATVMFDANILAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILAGSSFMSHA
KKVNEMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTLEAVDILKLMTSTYIVALCQAVDLRHLEENIKSSVKNCVTQVAKKVL

TMNPTGDLSSARFSEKNLLTAIDREAVFSYADDPCSANYPLMQKLRAVLVEHALTSGDAEPEASVFSKITKFEEELRSALPREIEAARVAVANGTAPVANRIVESRS
FPLYRFVREELGCVFLTGEKLKSPGEECNKVFLGISQGKLIDPMLDCLKEWNGEPLPIN

>K3Y5M0_Setaria_italica

MASNTAILESDPLNWGKAAAEMAGSHLDEVKRMVAQFREPLVKIEGSSLRVGQVAAVAAAKDASGVAVELDEDARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLINTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALGAMVCFDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILEGSSFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYMVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPVGELSSARFSEKDIITAIDREGVFTYAEDAASASLPLMQKLRAVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIKDSRS
FPVYRFVREELGCVFLTGEKLKSPGEECNKVFIGISQGKLIDPMLECLKEWDGKPLPIC

>K3YQG1_Setaria_italica

MAGNGLIVENDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVASARDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVTRAAMLVRINTLLQGYSGIRFEILEAITKLLNSGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTA
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVLYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILEGSEFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYMVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPTGDLSSARFSEKDLITAIDREGVFTYAEDAASASLPLMQKLRAVLVDHAFSSGDEPSMFSKITKFEEELRAVLPQEVEAARVAVAEGTAAVENRIKDSRSFPL
YRFVREELGCVFLTGEKLKSPGEECNKVFIGISQGKLIDPMLECLKEWDGKPLPIN

>K3YGI0_Setaria_italica

MAGNGLIVENDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGAGSDGHTLPSEVTRAAMLVRINTLLQGYSGIRFEILEAITKLLNSGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFRLNPKEGLAIVNGTSVGSALAATVLYDANLLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILEGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPAGELSSARFSEKDLLTAIDREGVFTYAEDAASASLPLMQKLRAVLVDHAFSSGDEPSVFSKITKFEEELRAVLPQEVEAARVAVAEGTAAVENRIKDSRSFPL
YRFVREELGCVFLTGEKLKSPGEECNKVFVGISQGKLVDPMLECLKEWDGKPLPIN

>B8A046_Zea_mays

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVASAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVTRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSSFMKQA
KKLNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAVDLRHLEENIKASVKNTVTQVAKKVL
TMNPSGELSSARFSEKELISAIDREAVFTYAEDAASGSLPLMQKLRAVLVDHALSSGDAEREPSVFSKITRFEEELRAVLPQEVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGERLKSPGEECNKVFVGISQGKLVDPMLECLKEWDGKPLPINVK

>I1IZQ0_Brachypodium_distachyon

MEYENGHAATYGDGLCVAAPLAPRADPLNWGKAAEELSGSHLDAVKRMVEEYRRPVVKMEGASLTIAQVAAVAAGAEARVELDESARGRVKESSDWVMNSMMNGTDS
YGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGEDGHVLPAAATRAAMLVRVNTLLQGYSGIRFEILETIATLLNANVTPCLPLRGTITASGDLVPLSYIAGL
VTGRPNSVATAPDGRKVNAAEAFKIAGIQHGFFELQPKEGLAMVNGTAVGSGLASMVLFEANILGVLAEVLSAVFCEVMNGKPEFTDHLTHKLKHHPGQIEAAAIME
HILEGSSYMMLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSE
LVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDILKLMSSTFLVALCQAIDLRHLEENVRSAV
KNCVTTVARKTLSTNVNGHLHNARFCEKDLLLTIDREAVFAYADDPCSANYPLMQKMRAVLVEHALANGEAERDVETSVFAKLAAFEQELRAVLPKEVEAARAAVEN
GTATKQNRIAECRSYPLYRFVREELGTEYLTGEKTRSPGEEVDKVFVAMNQGKHIDALLECLKEWNGEPLPLC

>D3JYP7_Bambusa_oldhamii

MECENGHVAANVSDLCMAKPPRADPLNWGKAAEELSGSHLDAVKRMVDEYRRPVVRIEGASLTIAQVAAVAAAGEARVELDESARGRVKESSDWVMNSMMNGTDSYG
VTTGFGATSHRRTKEGGALQRELIRFLNAGALDTGDDGHVLPAAATRAAMLVRINTLLQGYSGIRFEILETIAALLNANVTPCLPLRGTITASGDLVPLSYIAGLVT
GRPNSVAVAPDGRKVNAADAFKIAGIQHGFFELQPKEGLAMVNGTAVGSGPASMVLFEANILGVLAEVLSAVFCEVMNGKPEYTDHLTHKLKHYPGQIEAAAIMEHT
LEGSSYMMLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELV
NDLYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTGEATDILKLMSSTFLIALCQAIDLRHLEENVKNAVKN
CVTTVARKTLSTSATGDLHNARFCEKDLLKAIDREAVFAYADDPCSANYPLMQKMRAVLVEHALANGEAESNVDTSVFAKVATFEEELRAMLPREVEAARAAVENGT
AAKQNGITECRSYPLYRFVREELGTEYLTGEKARSPGEEVNKVFVALNQGKHIDALLECLKEWNGEPLPIC

>I1IBR6_Brachypodium_distachyon

MECENGLVGSLNGEGLCMSAPPRAAADPLNWAKTAEELAGSHLEEVKKMVAQFRMPLVKIEGATLGIAQVAAVAAGAGEARVELDESARGRVKESSDWVMNSMMNGT
DSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAITKLLNANVTPCLPLRGTITASGDLVPLSYIA
GLITGRQNSVAVAPDGSKVSAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANIQAILAEVLSAVFCEVMTGKPEFTDHLTHKLKHHPGQIEAAAI
MEHILEGSSYMKEAKKQGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKAIHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQF
SELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDILKIMSSTFLIALCQAIDLRHLEENMKT
AVRNCVMQVAKKTLSMNAMGGLHIARFCEKDLLTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDGERALETSIFAKVAEFEQNLRAALPKEVEAARASV
ENGTPLAPNRIKDCRSYPLYRFVREELGTEYLTGEKTRSPGEELNKVLVAMNQRKHIDPLLECLKEWNGEPLPLC

>B4FW68_Zea_mays

MESEAGLLVRSSLNGEGLCMPAPRADPLNWGKAAEGLSGSHLDEVKRMVAEFRDPLVKIQGASLSVAQVAAVAVGAGGGEARVELDESARERVRASSDWVMGSMMNG
TDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRVNTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYI
AGLITGRQNSVAVDPDGRKVGAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLAVLAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEAAA
VMEHILEGSSYMKLAKRLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQ
FSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDILKLMSSTFLIALCQAIDLRHLEENVK
AAVKNCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYALMQKLRAVLVEHALANGDAERDVDTSIFAKVAEFEQQVRAALPKEVEAARAA
VENGSPLVPNRIKECRSYPLYRFVREEVGTKYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKEWNGEPLPLC

>B6U0Z0_Zea_mays

MESEAGLLVRSSLNGEGLCMPAPRADPLNWGKAAEGLSGSHLDEVKRMVAEFRDPLVKIQGASLSVAQVAAVAVGAGGGEARVELDESARERVRASSDWVMGSMMNG
TDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRVNTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYI
AGLITGRQNSMAVAPDGRKVGAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLAVLAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEAAA
VMEHILEGSSYMKLAKRLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQ
FSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGDPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDILKLMSSTFLIALCQAIDLRHLEENVK
AAVKNCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYALMQKLRAVLVEHALANGDAERDVDTSIFAKVAEFEQQVRAALPKEVEAARAA
VENGSPLVPNRIKECRSYPLYRFVREEVGTEYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKEWNGEPLPLC

>K3YQC4_Setaria_italica

MECETGLVRSLHGDGLCMSAQAAAPRADPLNWGKAAEDLSGSHLDEVKRMVAEFREPLVKIQGASLSIAQVAAVAAGAGEARVELDESARERVKASSDWVMNSMMNG
TDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRINTLLQGYSGIRFEILETIAKLLNANVTPCLPLRGTITASGDLVPLSYI
AGLITGRQNSVAVAPDGRKVDAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLAIMAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEAAA
IMEHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQ
FSELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMTSTFLIALCQAIDLRHLEENMK
AAVKNCVTQVAKKTLSMNAMGGLHIARFCEKDLQTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDAERVVETSIFAKVAEFEQQVRAALPKEVEAARAA
VESGRPMVPNRIKECRSYPLYRFVREELGAEYLTGEKTRSPGEELNKVLVAINERKHIDPLLECLKEWNGEPLPLC

>C0LL35_Bambusa_oldhamii

MECENGQVASNGNGLCMATPRADPLNWGKAAEELMGSHLEEVKRMVAEYRQPVVKIEGASLRIAQVAAVAAGAGEAKVQLDDSARGRVKESSDWVMNSMMNGTDSYG
VTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGSDGHVLAAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAGLVT
GRENSVAVAPDGRKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHI
LEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLAIAAVGKLMFAQFSELV
NDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDILKIMSSTFLVALCQAIDLRHIEENVKSAVKS
CVMTVAKKTLSTNSTGDLHVARFCEKDLLKEIDREAVFAYADDPCSPNYPLMKKMRNVLVERALANGMAEFNAETSVFAKVAQFEEELRATLPRAVEAARAAVENGT
AATPNRITECRSYPLYRFVREELGTAYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIN

>A2X7F7_Oryza_sativa_indica

MECENGRVSANGMSGLCVAAPRADPLNWGKATEEMTGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAAGEARVELDESARERVKASSDWVMNSMMNGTDSYG
VTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAGLVT
GRENAVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHI
LEGSSYMKHAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELV
NDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDILKLMSSTFLIALCQAVDLRHIEENVKSAVKS
CVMTVAKKTLSTNSTGDLHVARFCEKDLLKEIDREAVFAYADDPCSHNYPLMKKLRNVLVERALANGAAEFNADTSVFAKVAQFEEELRATLPGAIEAARAAVENGT
AAIPSRITECRSYPLYRFVREELGTKYLTGEKTRSPGEELNKVLVAINEGKHIDPLLECLKEWNGEPLPIC

>C0HJ40_Zea_mays

MECENGRGVAATNSDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLSIAQVAAVATGAGEARVELDESARSRVKASSDWVMTSMMNGTDS
YGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAETTRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGL
VTGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIME
HILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSE
LVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAV
KSCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFNAETSVFAKVAQFEEDLRAALPKAVEAARAAVEN
GTAGIPNRIAECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>A0A096TA22_Zea_mays

MACDSPCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLSIAQVAAVATGVGEARVELDESARSRVKASSDWVMSSMMNGTDSYGVTTGFGATS
HRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLVTGRENSVAVA
PDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHILEGSSYMKL
AKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDYYNNGLP
SNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNSVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKIMSSTFLIALCQAVDLRHIEENVKSAVKSCVMTVAKKT
LSTDSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFNAETSVFAKVAQFEEDLRAALPKAVEAARAAVENGTAAIPNRITD
CRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>I1IBR8_Brachypodium_distachyon

MECENGQFAANGTGLCMATPRADPLNWGKAAEELTGSHLDEVKRMVAEYRKPVVTIEGATLSIAKVAAVAAAGEAKVELDESARERVKASSDWVMNSMANGVDSYGV
TTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGSDGHVLPAGATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAGLVTG
RENSVAVTPDGRKVNAAEAFKIAGIHGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAVLAEVISAVFCEVMNGKPEFTDHLTHKLKHHPGQIEAAAIMEHIL
EGSSYMKLAKKLGDLDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKAIHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQFSELVN
DFYNNGLPSNLSGSRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVEILKLMSSTFLVALCQAIDLRHIEENVKSAVKSC
VMTVAKRTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRGVLVERALSNGKAEFNAETSVFAKVAQFEEELRTALPKAVEAARSAVESGTA
ATPNRIKECRSYPLYRFVREELGTAYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>I1IBL7_Brachypodium_distachyon

MECENGQFAANGTGLCMATPSADPLNWGKAAEELTGSHLDEVKRMVAEYRKPVVTIEGATLSIAKVAAVAAAGEAKVELDESARERIKASSDWVMNSMMNGTDSYGV
TTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGSDGHVLPAGATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAGLVTG
RENSVAVTPDGRKVNAAEAFKLAGIHGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAVLAEVISAVFCEVMNGKPEFTDHLTHKLKHHPGQIEAAAIMEHIL
EGSSYMKLAKKLGDLDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKAIHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQFSELVN
DFYNNGLPSNLSGSRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVEILKLMSSTFLVALCQAIDLRHIEENVKIAVKSC
VMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRSVLVERALSNGMAEFNAETSVFAKVAQFEEELRTVLPKAVEAARAAVESGTA
ATPNRIKECRSYPLYRFVREELGTAYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>I1IBR7_Brachypodium_distachyon

MARENARVAAANGICTAIQHADPLNWGKAAEELTGSHLDEVKRMVVEYREPVVTIEGASLSIAKVAAVAAAGEAKVELDESARERVKASSDWVMNSMANGVDSYGVT
TGFGATSHRRTKEGGALQRELIRFLNAGAFGTGSDGHVLPAGATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAGLVTGR
ENSVAVAPDGRKVNAAEAFKIAGIHGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAVLAEVISAVFCEVMNGKPEFTDHLTHKLKHHPGQIEAAAIMEHILE
GSSYMKLAKKQGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKAIHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQFSELVND
LYNNGLPSNLSGGRNPSLDYGLKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLVALCQAIDLRHIEENVKSAVKSCV
MTVAKKTLSTNSTGDLHVSRFCEKDMLQEIDREAVFAYADDPCNHNYPLMKKLRGVLVESALANGVAEYNVETSVFAKVAQFEEELRAALPKAVEAARAAVESGTAA
TPNRITECRSYPLYRFVREELGTVFLTGEKTRSPGEELNKVLMAINQGKHIDPLLECLKEWNGEPLPIC

>A0A0Q3JNS6_Brachypodium_distachyon

MTMASKNVHVSADGYLILCPATSQHADPLNWGKAAEALTGSHLEEVKRMVAEYRQPVVTIEGASLSIAKVAAVAAAGEAQVQLDESARERVKASSDWVMDSMANGVD
SYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGSDGHVLPAGATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAG
LVTGRENSVAVAPDGSKVNAAEAFKIAGIHGGFFELQPKEGLAMVNGTAVGSGLASTVLFDANVLAVMAEVISAVFCEVMNGKPEFTDHLTHKLKHHPGQIEAAAIM
EHILEGSSYMKLAKKLGDLDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKAIHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQFS
ELVNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISARKTAEAIEILKLMTSTFLVALCQAIDLRHIEENVKSA
VTSCVRAVAKKTLSTNSAGGLHVARFSEKDLIQEIDREAVFAYADDPCNPNYPLMKKLRGVLVERALANGVAEFDAETSVFAKVARFEEELRAALPVAVEAARAAVE
SGTAEAPNRIAECRSYPLYRFVRQELGTVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>I1IAZ3_BRADI_Brachypodium_distachyon

MACENGQVAANGICTAIQHADPLNWGKAAEALTGSHLEEVKRMVAEYRQPVVTIEGASLSIAKVAAVAAAGEAQVQLDESARERVKASSDWVMDSMANGVDSYGVTT
GFGATSHRRTKEGGALQRELIRFLNAGAFGTGSDGHVLPAGATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYIAGLVTGRE
NSVAVAPDGSKVNAAEAFKIAGIHGGFFELQPKEGLAMVNGTAVGSGLASTVLFDANVLAVMAEVISAVFCEVMNGKPEFTDHLTHKLKHHPGQIEAAAIMEHILEG
SSYMKLAKKLGDLDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKAIHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQFSELVNDF
YNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISARKTAEAIEILKLMTSTFLVALCQAIDLRHIEENVKSAVTSCVR
AVAKKTLSTNSAGGLHVARFSEKDLIQEIDREAVFAYADDPCNPNYPLMKKLRGVLVERALANGVAEFDAETSVFAKVARFEEELRAALPVAVEAARAAVESGTAEA
PNRIAECRSYPLYRFVRQELGTVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>A0A059Q1B2_Saccharum_hybrid_cultivar_R570

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEAEKDASASVFSKINRFEETLREVLPREMEAARVAFETGTAPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC

>U3M000_Saccharum_hybrid_cultivar_Co_93009

GKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHAKKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQG
TPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTSNLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILK
LMSSTYIVALCQAIDLR

>M4XZQ1_Saccharum_hybrid_cultivar_CP69-1062

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQRYSGIPFEILEAITKLLNTGVSPCLPLRGTITASRDLVPLSYIAGLITGRPNAQATTV
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGSSVGSALAATVMYDAYVLTVLSEVLSAVLCEVMNRMPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLIKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVMDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGLKGTENAMASYCSELQYLGNPITNHVQSAEQLNQDVNSFGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNRVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYSEDPASGSLPLMQKLRSVLVDHALSSGDAGTGPLRVLQDHQIRGGAPRGAGPGGGRRPASPWAEGTAPGRNRNWDSR
SFPLYRFVREELGCVFVTGEKLKSPGEECSKVFNGISQGKLVDPKLECLKEWDGKPLPINVVN

>W5RSK7_Saccharum_hybrid_cultivar_HSF-240

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGTDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEDITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSPLAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKTSVKNTVTQVAKKVL
TMNPSGELSSARFSEKELISAIEREDVCTHAEDPASVSLPLMQKLRTVLVDHALSSSDAGTGALRVLQDHHVRGGAPRGAAPGGGRRPALPWAEGTAPGRNRSWDSR
SFPLYRFVREELGCVFLTGEKLKSPGEECTKVFPGISQGKLVDPMLECLKEWDGKPLPIN

>A2IBN5_Saccharum_officinarum

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTV
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLTVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAGTGALRVLQDHQFRGGAPRGAGPGGGRRPASPWAEGTAPGRNRNWDSR
SFPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGISQGKLVDPMLECLKEWDGKPLPINVVN

>W5RSQ2_Saccharum_hybrid_cultivar_Co1148

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAATLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYVAGLITGRPNAQATTI
GGRKVDAAGAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVLGAEQQDQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELISAIDREAVFTYAEDPASGSLPLMQKLRPVLVDHALSSGDAGTGALRVLQDHQVRGGAPRGAGPGGGRRPALPWAEGTAPGANRTWDSR
SFPLYRFVREELGCVFLTGEKLKSPGEECNKVFPGISQGKLVDPMLECLKEWDGKPLPINVVNC

>M1MQ13_Saccharum hybrid cultivar ROC22 PE=2 SV=1

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGAITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAILSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAEREPSVFSKITRFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGISQGKLVDPMLECLKEWDGKPLPINIVN

>C5XXU0_Sorghum_bicolor

MACENGRVAATNGDGLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGTKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
SCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVRQEVGAVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>C5XXT9_Sorghum_bicolor

MECETGLVRSLNGDGLCMSAQAAPRGADPLNWGKAAEDLSGSHLEEVKRMVAEFRDPVVKIQGASLSIAQVAAVAAGAGEARVELDESARERVKASSDWVMSSMMNG
TDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLSYI
AGLITGRQNSVAVAPDGRKVDAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLAIMAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEAAA
IMEHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMFAQ
FSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIDILKLMTSTFLIALCQAIDLRHIEENVK
AAVKNCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDAERVVETSIFAKVAEFEQQVRAALPKEVEAARAA
VESGNPLVPNRIKECRSYPLYRFVREEVGTQYLTGEKTRSPGEELNKVLVAINQRKHVDPLLECLKEWNGEPLPLC

>C5XXT8_Sorghum_bicolor

MAGNGAIVESDPLNWGAAAAELSGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVASAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTV
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAIFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVLRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASASLPLMTKLRAVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGINQGKLVDPMLECLKEWDGKPLPINVVN

>C5YCD6_Sorghum_bicolor

MASNTAILESDPLSWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIQGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNCVMAAARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALASGEAEKDASASVFSKINRFEEALREALPREMEAARVAFETGTAPIGNRIKDS
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC

>C5XXU3_Sorghum_bicolor

MACDNGRVAATNGDGLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGTKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
SCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVRQEVGAVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC

>evm.model.scga7_uti_cns_0033817.4

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTV

DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAILSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGISQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_uti_cns_0037780.4

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGISQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_unitig_176128.1

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGISQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_uti_cns_0098816.2

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLTVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRAVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGINQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_uti_cns_0130599.1

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGISQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_uti_cns_0122071.2

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRAVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFIREELGCVFLTGEKLKSPGEECTKVFNGINQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_uti_cns_0044626.3

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHIEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRAVLVDHALSSGDAEREPSVFSKITKFEEELRAVLPREVEAARVAVAEGTAPVANRISDSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGINQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_uti_cns_0064152.1

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLTVLSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGSLPLMQKLRSVLVDHALSSGDAEREPSVFSKITRFEEELRAVLPREVEAARVAVAEGTAPVANRIADSRS
FPLYRFVREELGCVFLTGEKLKSPGEECTKVFNGISQGKLVDPMLECLKEWDGKPLPINVVN*

>evm.model.scga7_uti_cns_0252506.1

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAVVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEAEKDASASVFSKINRFEETLREVLPREMEAARVAFETGTAPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC*

>evm.model.scga7_uti_cns_0004769.3

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTTLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEADKDASASVFSKINRFEETLREALPREMEAARVAFETGTAPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC*

>SHCRBa_019_F13_F_30_1

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL

TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEAEKDASASVFSKINRFEETLREVLPREMEAARVAFETGTAPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC

>evm.model.scga7_uti_cns_0313854.1

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPIVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAVVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEAEKDASASVFSKINRFEETLREVLPREMEAARVAFETGTAPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC*

>evm.model.scga7_unitig_244058.1

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDAAGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSEVQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEADKDASASVFSKINRFEETLREALPREMEAARVAFETGTTPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC*

>evm.model.scga7_uti_cns_0185632.2

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPVVKIEGSTLRVGQVAAVAAAKDAAGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSEVQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEADKDASASVFSKINRFEETLREALPREMEAARVAFETGTTPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC*

>evm.model.scga7_uti_cns_0252674.1

MASNTAILESDPLNWGKAAAELTGSHLDEVKRMVAQFRDPIVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGNGSDGHTLPSEVSRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQAVTV
DGRKVDAAEAFKVAGIEGGFFKLNPKEGLAIVNGTSVGSALAAMVCFDANVLAVLSSVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIESAAIMEHILDGSSFMKHA
KEVNAMDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLANPITNHVQSAEQHNQDVNSLGLVSARKTAEAVDILKLMSSTYMVALCQAVDLRHLEENLKSAVKNSVMAVARKVL
TTSLDGDLHSARFSEKALLTAIDREAVYGYYDDPCSANSPLMKKIRAVLVDHALANGEAEKDASASVFSKINRFEETLGEVLPREMETARVAFETGTAPIANRIKES
RSYPLYRFIRQDLGAVYLTGEKLKSPGEECNKVFLALSEGKLIDPMLECLKEWDGKPLPIC*

>evm.model.scga7_uti_cns_0304192.1

MECENGHVAAASGNGVCLATPREADPLNWGKAAEDLTGSHLEAVKRMVEEYRRPLVKIEGGSLTVAQVAAVAAAGEARVELDESARGRVKASSDWVMDSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGDDGHVLPAAATRAAMLVRINTLLQGYSGIRFEILETIAALLNANVTPCLPLRGTITASGDLVPLSYIAGLV
TGRPNSTAVAPDGRKVDAAEAFKIAGIQHGFFELQPKEGLAMVNGTAVGSGLASIVLFEANVLAVLAEVMSAVFCEVMNGKPEYTDHLTHKLKHHPGQIESAAIMEH
ILDGSSYMMLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDVLKLMSSTFLIALCQAVDLRHLEENLKSAVK
SCVMTVAKKTLSTSATGTLHNARFCEKDLLTAIDREAVFAYADDPCSANYPLMQKMRSVLVEHALANGDAERNPDTSVFAKLATFEEELRAALPREVEAARAAVENG
TAAIPNRIAECRSYPLYRFVREELGTEYLTGEKTRSPGEEVNKVFVAMNLGKHIDAVLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0000426.9

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGGARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIETLKLMSSTFLIALCQAVDLRHIEENVKSAVK
SCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIDREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAVIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0341334.1

MECDNGHVAAASGNGVCLATPRAADPLNWGKAAEDLTGSHLEAVKRMVEEYRRPLVKIEGASLTVALVAAVAAAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGDDGHVLPAAATRAAMLVRINTLLQGYSGIRFEILETIAVLLNANVTPCLPLRGTITASGDLVPLSYIAGLV
TGRPNSTAVAPDGRKVDAAEAFKIAGIQHGFFELQPKEGLAMVNGTAVGSGLASIVLFEANVLAVLAEVMSAVFCEVMNGKPEYTDHLTHKLKHHPGQIESAAIMEH
ILDGSSYMMLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDFYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDVLKLMSSTFLIALCQAVDLRHLEENLKSAVK
SCVMTVAKKTLSTSATGTLHNARFCEKDLLTAIDREAVFAYADDPCSANYPLMQKMRSVLVEHALANGEAERNPDTSVFAKVATFEEELRAALPREVEAARAAVENG
TAAIPNRIAECRSYPLYRFVREELGTEYLTGEKTRSPGEEVNKVFVAMNLGKHIDAVLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0106227.1

MECETGFVRSLNGDGLCMSASAAAPRATDPLNWGKAAEDLSGSHLEEVKRMVAEFREPVVKIQGASLSIAQVAAVAAGAGSEARVELDESARERVKASSDWVMSSMM
NGTDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLS
YIAGLITGRQNSVAVAPDGRKVDAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLATMAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEA
AAIMEHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMF
AQFSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMTSTFLIALCQAIDLRHIEEN
VKAAVKNCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDAERVVETSIFAKVAEFEQQVRAALPKEVDAAR
AAVESGNPLVPNRIKECRSYPLYRFVREEVGTEYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKEWNGEPLPLC*

>evm.model.scga7_uti_cns_0043984.2

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
SCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIDREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0055007.2

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
GCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0116868.1

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKMKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKTAVK
SCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIDREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0194663.1

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKMKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKTAVK
SCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIDREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0034023.1

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
GCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAIN

QGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0084431.3

MACENSNVTANGDGLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDEYARGRVKASSDWVMNSMMNGTDSYG
VTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLVT
GRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAVLAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHI
LEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELV
NDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVKS
CVMMVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMNKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARTAVENGT
AAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0064033.1

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
GCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFFKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0129632.1

MECENGRVAATNGDSLCMATPRTDPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
SCVMTVAKKTLSTDSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0351321.1

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAVLAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
SCVMTVAKKTLSTDSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0129758.8

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV
TGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKNAVK
GCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIDREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPDEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0111854.1

MECETGFVRSLNGDGLCMSAPAPAPAPRASDPLNWGKAAEDLSGSHLEEVKRMVAEFREPVVKIQGASLSIAQVAAVAAGAGGEARVELDESARERVKASSDWVMSS
MMNGTDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVP
LSYIAGLITGRQNSVAVAPNGRKVDAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLATMSEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQI
EAAAIMEHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKL
MFAQFSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMTSTFLIALCQAIDLRHIE
ENVKAAVKNCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDAERVVETSIFAKVAEFEQQVRAALPKEVDA
ARAAVESGNPLVPNRIKECRSYPLYRFVREEVGTEYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKEWNGEPLPLC*

>evm.model.scga7_uti_cns_0171079.1

MACENSNVTANGDGLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSYG
VTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLVT
GRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHI
LEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELV
NDYYNIGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSTVKS
CVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENGT
AAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0081048.2

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMTNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLV

TGRENSVAVAPDGSKVNAAQAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAADIMEH
ILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSEL
VNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVK
GCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENG
TAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0033817.5

MECETGFVRSQNGDGLCMSAAAPAPRASDPLNWGKAAEDLSGSHLEEVKRMVAEFREPVVKIQGASLSIAQVAAVAAGAGSEARVELDESARERVKASSDWVMSSMM
NGTDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLSAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLS
YIAGLITGRQNSVAVAPDGRKVDAAEAFKIAGIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLATMAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEA
AAIMEHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMF
AQFSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMTSTFLIALCQAIDLRHIEEN
VKAAVKNCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDAERVVETSIFAKVAEFEQQVRAALPKEMDAAR
AAVESGNPLVPNRIKECRSYPLYRFVREEVGTEYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKEWNGEPLPLC*

>evm.model.scga7_uti_cns_0204213.1

MECETGFVRSQNGDGLCMSAAAPAPRASDPLNWGKAAEDLSGSHLEEVKRMVAEFREPVVKIQGASLSIAQVAAVAAGAGSEARVELDESARERVKASSDWVMSSMM
NGTDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGADGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIAKLLNANVTPCLPLRGTITASGDLVPLS
YIAGLITGRQNSVALAPDGRKVDAAEAFKIASIEHGFFELQPKEGLAMVNGTAVGSGLASTVLFEANVLATMAEVISAVFCEVMTGKPEFTDHLTHKLKHHPGQIEA
AAIMEHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRGKALHGGNFQGTPIGVSMDNTRLALAAIGKLMF
AQFSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAVDILKLMTSTFLIALCQAIDLRHIEEN
VKAAVKNCVTQVAKKSLSLNARGGLHNARFCEKDLQTAIDREAVFAYADDPCSPNYPLMQKLRAVLIEHALANGDAERVVETSIFAKVAEFEQQVRAALPKEMDAAR
AAVESGNPLVPNRIKECRSYPLYRFVREEVGTEYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKEWNGEPLPLC*

>evm.model.scga7_uti_cns_0033817.6

MACENSNVTANGDGLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARDRVKASSDWVMNSMMNGTDSYG
VTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLVT
GRENSVAVAPDGSKVNAAEAFKIAGILGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHI
LEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELV
NDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVYLRHIEENVKSAVKS
CVMTVAKKTLSTNSTGGLHVARLCEKDLLQEIDREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARTAVENGT
AAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLVAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0055321.1

MECENGRVAATNGDSLCMATPRADPLNWGKAAEELMGSHLDEVKRMVSEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSY
GVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAKATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAG
LVTGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIM
EHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFS
ELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSA
VKGCVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVE
NGTAAIPNRITECRSYPLYRFVREELGAVYLTGEKTRSPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0407956.1

MACENSNVTANGDGLCMATPRADPLNWGKAAEELMGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSYG
VTTGFGATSHRRTKEGGALQRELIRFLNAGAFGTGTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLVT
GRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQPKEGLAMVNGTAVGSGLASTVLFEANILAVLAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHI
LEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELV
NDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASYCSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVKS
CVMTVAKKTLSTNSTGGLHVARFCEKDLLQEIEREAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFDAETSVFAKVAQFEEELRAALPKAVEAARAAVENGT
AAIPNRITECRSYPLYRFVREEVGTEYLTGEKTRSPGEELNKVLVAINQRKHIDPLLECLKGWNGEPLPIC*

>evm.model.scga7_uti_cns_0084431.4

MGSHLDEVKRMVAEYRQPLVKIEGASLRIAQVAAVAAGAGEARVELDESARGRVKASSDWVMNSMMNGTDSYGVTTGFGATSHRRTKEGGALQRELIRFLNAGAFGT
GTDGHVLPAEATRAAMLVRINTLLQGYSGIRFEILEAIVKLLNANVTPCLPLRGTVTASGDLVPLSYIAGLVTGRENSVAVAPDGSKVNAAEAFKIAGIQGGFFELQ
PKEGLAMVNGTAVGSGLASTVLFEANILAILAEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGQIEAAAIMEHILEGSSYMKLAKKLGELDPLMKPKQDRYALRTSPQ
WLGPQIEVIRAATKSIEREINSVNDNPLIDVARSKALHGGNFQGTPIGVSMDNTRLAIAAIGKLMFAQFSELVNDYYNNGLPSNLSGGRNPSLDYGFKGAEIAMASY
CSELQFLGNPVTNHVQSAEQHNQDVNSLGLISSRKTAEAIEILKLMSSTFLIALCQAVDLRHIEENVKSAVKSCVMTVAKKTLSTDSTGGLHVARFCEKDLLQEIER
EAVFAYADDPCSANYPLMKKLRNVLVERALANGAAEFNAETSVFAKVAQFEEELRAALPKAVEAARTAVENGTAAIPNRITECRSYPLYRFVREEVGAVYLTGEKTR
SPGEELNKVLLAINQGKHIDPLLECLKEWNGEPLPIC*

>evm.model.scga7_uti_cns_0051921.1

MAGNGAIVESDPLNWGAAAAELAGSHLDEVKRMVAQARQPVVKIEGSTLRVGQVAAVAAAKDASGVAVELDEEARPRVKASSEWILDCIAHGGDIYGVTTGFGGTSH
RRTKDGPALQVELLRHLNAGIFGTGSDGHTLPSEVVRAAMLVRINTLLQGYSGIRFEILEAITKLLNTGVSPCLPLRGTITASGDLVPLSYIAGLITGRPNAQATTI
DGRKVDAAEAFKIAGIEGGFFKLNPKEGLAIVNGTSVGSALAATVMYDANVLAILSEVLSAVFCEVMNGKPEYTDHLTHKLKHHPGSIEAAAIMEHILDGSAFMKHA
KKVNELDPLLKPKQDRYALRTSPQWLGPQIEVIRAATKSIEREVNSVNDNPVIDVHRGKALHGGNFQGTPIGVSMDNARLAIANIGKLMFAQFSELVNEFYNNGLTS
NLAGSRNPSLDYGFKGTEIAMASYCSELQYLGNPITNHVQSAEQHNQDVNSLGLVSARKTAEAIDILKLMSSTYIVALCQAIDLRHLEENIKTSVKNTVTQVAKKVL
TMNPSGDLSSARFSEKELITAIDREGVFTYAEDPASGC*

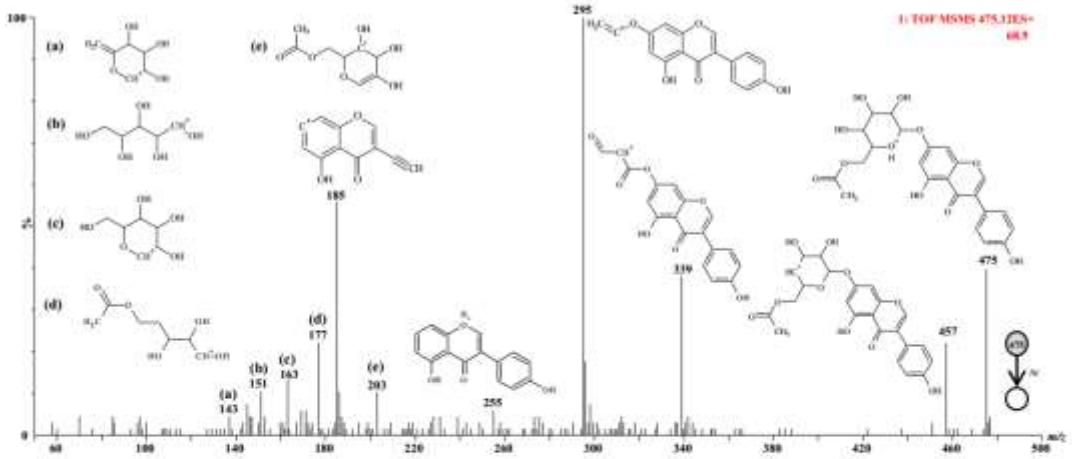
APPENDIX B. Sugarcane meristem metabolites identified using GC-TOF-MS approach. RI: Retention Index. Mass: Characteristic masses of each metabolite. QuantMass: mass used for quantification of each metabolite. t-test was used to determine significative changes (p-value < 0.05) between infected and control samples of the same age. Relativized medians of each metabolite in 5, 65, 100 and 120 DAI control and inoculated samples.

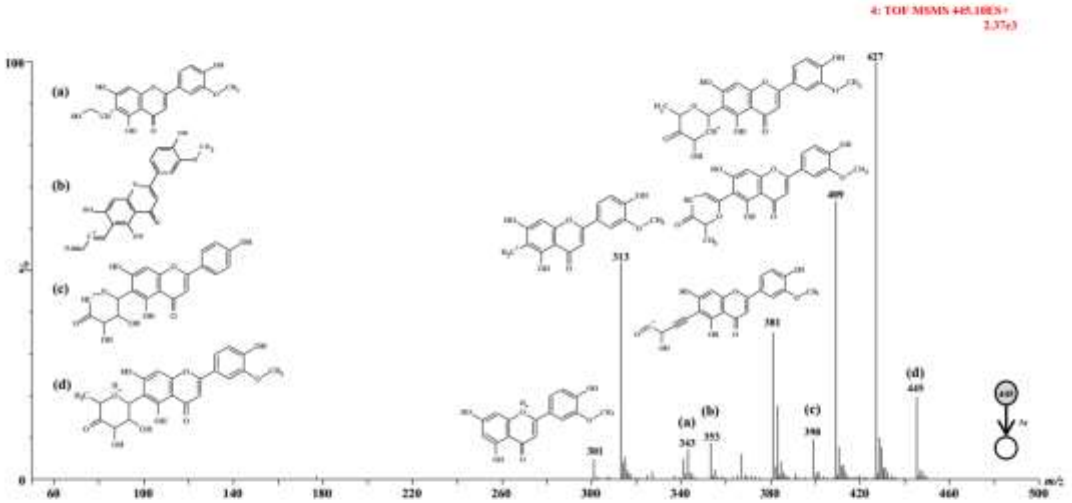
Metabolite ID	RI	MASS	quantMass	p-value (infected versus control)				Relativized medians							
				5 DAI	65 DAI	100 DAI	120 DAI	5DAI CONT	5DAI INOC	65 DAI CONT	65 DAI INOC	100 DAI CONT	100 DAI INOC	120 DAI CONT	120 DAI INOC
(r|x) [Fructose|Sorbose]	586290	217 103 307 189 277 129	217	0.394	0.089	0.645	0.157	2.30	1.72	0.82	0.50	0.98	0.79	0.65	1.05
(r|x) [Maltose|Laminaribiose]	870980	204 217 361 103 117 271	204	0.862	0.528	0.432	0.772	3.64	3.94	0.65	0.76	0.90	1.01	1.10	1.22
(r|x) [Raffinose|1-Kestose|Inulotriose]	1033310	217 361 103 129 169 271 204 437 451	217	0.009	0.658	0.854	0.024	11.91	41.87	0.90	0.58	0.55	0.26	2.67	0.70
(r|x) [Ribitol|Arabitol]	502990	117 217 319 129 157 243 103	117	0.083	0.004	0.018	0.258	2.38	3.77	0.14	0.51	0.43	0.91	0.41	1.02
(r|x) Cellobiose	862260	204 217 361 103 117 129 169	204	0.702	0.054	0.095	0.327	2.94	3.16	0.55	0.73	0.86	0.63	1.50	1.25
(r|x) Fructose	585640	217;307;103;189;129;277;364	217	0.070	0.119	0.810	0.920	1.90	0.83	1.22	0.31	0.56	0.76	0.82	0.99
(r|x) Putrescine	518310	174 100 86 214 200 130	174	0.181	0.992	0.037	0.075	1.10	1.52	0.88	0.83	1.12	0.74	1.14	1.68
(r|x) Shikimate	586300	204 255 133 282 372	204	0.343	0.382	NA	0.076	1.19	1.02	0.64	0.77	0.99	0.60	2.28	1.59
(r|x) Sorbose	587160	217;307;103;189;129;277;364	217	0.406	0.108	0.645	0.157	2.32	1.75	0.82	0.51	1.00	0.81	0.67	1.07
(r|x) Stearate	771680	117 132 129 145 341 201	117	0.101	0.029	0.423	0.114	1.60	1.89	0.47	0.81	0.87	0.71	1.23	0.94
(r|x) Talose	588520	160 319 157 205 217 129 103	160	0.336	0.055	0.125	0.226	4.91	6.13	0.22	0.50	0.47	0.87	1.19	0.91
(r|x) Xylitol	501050	217 117 103 319 307 243 129	217	0.253	0.141	0.098	NA	2.31	2.97	0.31	0.77	0.35	0.80	0.92	1.27
(r|z) Spermidine	724770	144 116 174 156 201 100	144	0.442	0.000	0.807	0.122	1.02	1.10	0.39	1.35	1.14	1.02	0.82	1.06
[Fructose|Psicose]	580060	103 217 117 307 133 189 364 277	103	0.091	0.089	0.393	0.847	2.18	1.06	1.14	0.42	0.50	0.86	0.61	1.15
[Fumarate|Maleate]	372020	245 143 133 217 115 155	245	0.651	0.055	0.045	0.362	2.14	2.25	0.28	0.51	1.10	0.64	1.33	1.07
[Glc 6-P|Gal 6-P]	773680	299 387 160 315 217 103 357	299	0.921	0.150	0.732	0.000	0.99	0.97	0.74	0.93	1.36	1.22	0.80	1.71
[N-Acetyl-[Glc|Gal]]	671810	205 319 117 129 103 157 229	205	0.189	0.666	0.306	0.937	6.16	9.10	0.68	0.63	1.40	0.81	0.80	0.96
[Sorbose|Tagatose]	583070	103;217;307;189;277;364	103	0.174	0.101	0.716	0.758	1.86	1.16	1.22	0.46	0.55	0.74	0.83	0.99
[Talose|Altrose|Gulose]	593550	205 217 129 319 103 160 117	205	0.768	NA	NA	NA	1.99	2.11	0.00	0.93	0.01	0.02	NA	0.69
[Xylose|Arabinose|Lyxose]	498180	103 217 189 307 277 233	103	0.114	0.295	0.979	0.159	3.41	4.87	0.55	0.70	0.93	0.89	1.08	0.78
Alanine	209220	116 218 100 190 133	116	0.551	0.021	0.253	0.910	1.05	0.88	2.73	1.19	0.63	0.81	1.14	0.99
Arginine	491110	142 102 128 162 204	142	0.665	0.003	0.009	0.034	0.10	0.11	2.43	1.10	1.34	3.46	0.63	1.32
Asparagine	550410	116 132 188 231 100 141	116	0.703	NA	0.478	0.062	0.47	0.99	1.77	1.38	7.87	1.65	1.00	0.39
Aspartate	458230	232 100 218 117 188 202	232	0.403	0.493	0.022	0.001	0.55	0.68	1.09	1.01	1.48	1.07	0.83	1.35
Benzoate	348140	105 135 179 194	105	0.190	0.048	0.361	0.097	1.62	2.04	0.44	0.83	0.90	0.70	1.45	0.96
Caffeate	745060	219 191 249 133 117 293	219	0.134	0.084	0.011	0.963	1.50	1.93	0.59	0.93	1.23	0.44	1.17	1.18
cis-Aconitate	582460	229 285 211 133 375 97	229	0.776	0.220	0.297	0.381	1.51	1.78	0.62	0.71	1.06	0.88	1.22	1.08

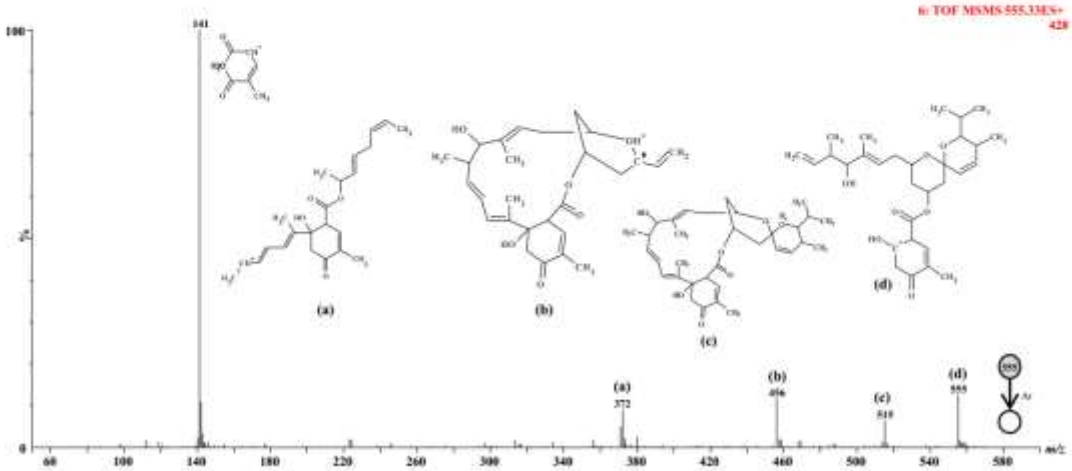
Metabolite ID	RI	MASS	quantMass	p-value (infected versus control)				Relativized medians							
				5 DAI	65 DAI	100 DAI	120 DAI	5DAI CONT	5DAI INOC	65 DAI CONT	65 DAI INOC	100 DAI CONT	100 DAI INOC	120 DAI CONT	120 DAI INOC
Citrate	593510	273 211 183 375 347 465	273	0.963	0.263	0.910	0.002	1.17	1.12	0.59	0.87	1.63	1.51	0.58	2.04
Citrulline	606380	157 142 256 100 115 218	157	0.571	0.425	0.031	0.174	0.53	0.71	1.23	1.43	0.81	2.09	0.74	1.07
Dehydroascorbate	625460	173 316 157 245 231 129	173	0.208	0.075	0.564	0.109	2.49	3.26	0.50	0.72	1.19	1.07	1.04	0.78
Deoxyribose	471410	103 217 133 117 173 307	103	0.531	NA	NA	NA	1.02	1.37	NA	NA	NA	NA	NA	NA
Galactinol	940240	204 217 129 103 305	204	0.749	0.141	0.008	0.006	1.01	1.04	1.47	1.09	1.73	0.61	0.58	0.32
Gluconate	627390	333 292 217 189 117 103 133	333	0.474	0.019	0.226	0.219	2.97	3.79	0.55	1.11	0.58	1.31	0.32	0.67
Gluconolactone	625920	220 129 319 243 229 157	220	0.176	0.034	0.123	0.196	1.62	2.06	0.28	0.85	0.95	0.68	1.07	0.83
Glucose_1	591390	205;160;103;129;319;217	205	0.074	0.126	0.788	0.967	2.31	1.09	1.21	0.30	0.51	0.76	0.54	0.71
Glucose_2	599440	103 205 160 129 117 217 319	103	0.172	0.085	0.425	0.098	2.03	1.23	0.98	0.31	0.97	0.70	0.61	1.38
Glutamate	508250	246 128 156 100 230 348	246	0.558	0.754	0.301	0.016	0.89	0.95	1.10	1.07	1.38	1.19	0.83	1.21
Glutamine_1	530570	227 156 203 128 317 139	227	0.839	0.838	0.365	0.846	0.10	0.10	0.93	0.94	1.32	1.12	1.17	1.11
Glutamine_2	598770	156;128;114;139;203	156	0.229	0.366	0.864	0.155	2.69	0.69	2.13	0.88	0.31	0.32	4.02	1.02
Glycerate	345490	189 103 292 205 133 307	189	0.025	0.002	0.290	0.134	0.94	1.12	0.54	0.94	1.07	0.94	0.98	1.18
Glycerol	293000	117 103 205 177 218 133	117	0.298	0.207	0.298	0.004	1.88	2.29	0.62	0.75	1.17	0.91	1.27	0.52
Glycine	325680	174 248 100 86 133 276	174	0.959	0.002	0.706	0.025	0.92	0.90	1.75	1.14	1.35	1.24	1.07	0.64
Histidine	678730	154 254 100 218 238	154	0.836	0.759	0.114	0.779	0.52	0.59	1.44	1.71	3.06	6.97	0.56	0.55
Isocitrate	598210	273 245 375 211 95	273	0.923	0.641	0.313	0.850	0.78	0.77	1.03	1.07	1.15	0.85	1.33	1.35
Isoleucine	319600	158 100 218 133 170 232	158	0.274	0.608	0.124	0.549	0.74	0.84	1.80	2.19	1.21	2.32	1.03	0.92
Itaconate	387270	215 259 133 97 117	215	0.163	0.160	0.155	0.956	1.63	1.90	0.56	0.83	0.91	0.70	1.15	1.15
Lactate	189530	117 191 219 133 101	117	0.203	0.711	0.523	0.078	1.86	2.58	0.73	0.79	0.65	0.72	2.59	1.01
Leucine	305880	158 102 232 218 260 100	158	0.673	0.599	0.279	0.803	0.68	0.71	1.73	1.44	0.99	1.30	0.94	0.92
Lysine	616600	174 156 100 128 230 317	174	NA	0.701	0.052	NA	NA	NA	0.90	0.87	0.92	1.69	0.99	NA
Lyxose	488070	103 217 189 117 160 307	103	0.001	0.023	0.168	0.018	0.64	1.00	1.67	0.95	2.31	3.10	0.72	0.26
Malate	442310	233 245 101 133 189 175 307 335	233	0.646	0.826	0.377	0.001	0.26	0.28	1.03	1.01	1.43	1.06	0.98	1.85
Methionine	474940	128 176 100 219 202 250 293	128	0.219	0.969	0.511	0.003	0.42	0.63	1.23	1.30	2.22	2.39	1.09	0.46
myo-Inositol	654460	217 191 305 318 103	217	0.065	0.016	0.118	0.362	1.51	1.89	0.71	1.06	1.15	0.86	0.69	0.52
N-Acetylgalactosamine	671810	205 319 117 129 103 157 229	205	0.106	0.108	0.483	0.302	1.86	2.74	0.52	1.00	1.02	0.79	0.93	0.67
Orthophosphate	333340	299 211 133 314 115 193	299	0.758	0.235	0.831	0.420	0.78	0.89	3.53	0.89	1.15	1.13	1.28	1.12
Phenylalanine	531550	218 192 100 266 130 294	218	0.797	0.461	0.399	0.168	0.43	0.45	1.12	1.31	1.07	0.93	1.38	1.13
Proline	339770	142 216 100 175 244 133	142	0.855	0.002	0.136	0.135	0.50	0.56	1.43	7.45	0.76	1.92	1.13	0.71
Pyruvate	222670	174 89 115 158 99	174	0.267	0.334	0.911	0.803	1.68	1.99	0.60	0.67	0.91	0.90	1.16	1.09
Quinate	578770	255 191 204 345 239 183	255	0.973	0.063	0.108	0.003	1.33	1.31	0.44	0.80	0.83	0.55	2.93	1.29

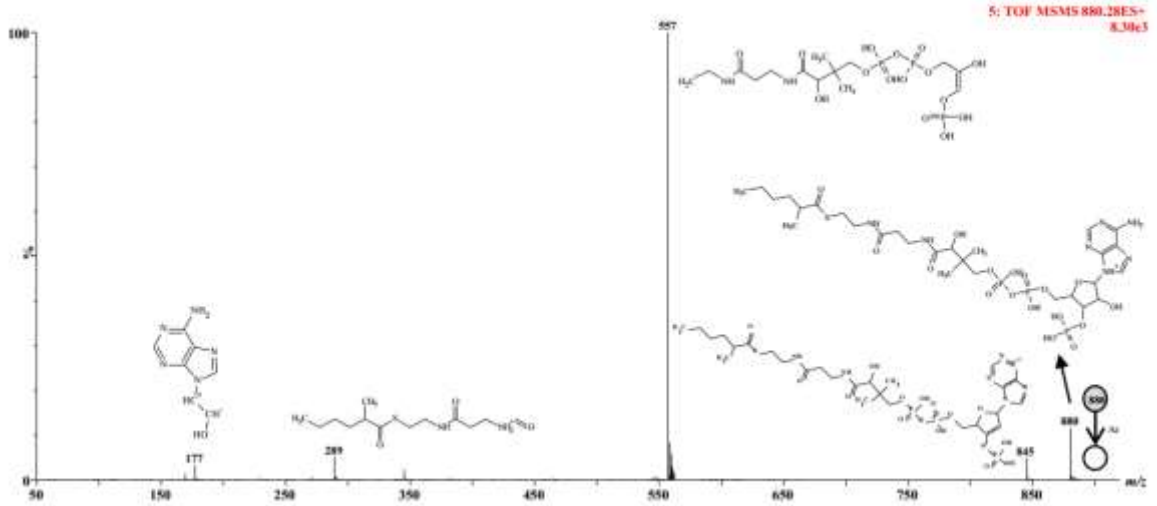
Metabolite ID	RI	MASS	quantMass	p-value (infected versus control)				Relativized medians							
				5 DAI	65 DAI	100 DAI	120 DAI	5DAI CONT	5DAI INOC	65 DAI CONT	65 DAI INOC	100 DAI CONT	100 DAI INOC	120 DAI CONT	120 DAI INOC
Quinic acid	578770	255 191 204 345 239 183	255	0.706	0.090	0.191	0.057	1.16	1.22	0.49	0.88	0.85	0.66	1.99	1.32
Rhamnose	516560	117 277 160 219 321 129	117	0.958	0.122	0.004	0.005	1.45	1.40	0.61	0.75	1.12	0.66	1.39	0.90
Ribonate	534600	217 189 103 133 292 333	217	0.212	0.001	0.040	0.698	1.10	1.34	0.50	1.03	1.09	0.85	1.03	0.96
Serine	358580	204 218 100 188 278 306	204	0.631	0.008	0.469	0.841	1.20	1.13	2.01	0.91	0.89	1.16	0.76	0.74
similar 2-Hydroxypyridine	204900	152 166 122 97 136	152	0.058	0.010	0.630	0.372	1.69	2.15	0.43	0.89	0.74	0.80	1.23	1.03
similar to Aminobutanoate	453640	174 304 246 216 100 86	174	0.548	0.028	0.655	0.737	0.80	0.92	2.24	1.26	1.08	0.93	0.88	0.81
Sucrose	841440	217 361 169 271 129 103 437 319	217	0.628	0.091	0.706	0.802	3.91	4.80	0.25	1.49	0.51	0.46	1.01	1.04
Tagatose	571210	103 307 217 189 277 364	103	0.262	0.762	0.113	0.016	1.01	0.64	0.85	0.87	1.05	1.61	0.92	1.51
Threonate	458980	292 220 205 117 189 319 103 130	292	0.485	0.049	0.287	0.008	0.48	0.55	1.18	0.98	1.44	1.28	0.79	1.08
Threonine	368400	219 117 291 101 129 320	219	0.796	0.920	0.102	0.121	0.48	0.50	1.80	1.82	1.28	2.48	0.70	0.90
Trehalose	571210	103 307 217 189 277 364	103	0.247	0.000	0.077	0.096	7.77	9.80	0.33	1.75	0.48	0.29	0.83	1.22
Tryptophan	791090	202 291 100 218 130	202	NA	NA	NA	NA	0.39	0.04	NA	NA	NA	NA	1.00	2.32
Tyrosine	659100	218 100 179 280 354 133	218	0.085	0.738	0.203	0.019	0.92	1.05	1.01	1.09	1.14	1.79	0.83	1.05
Valine	272610	144 218 100	144	0.705	0.334	0.420	0.100	0.69	0.73	2.05	1.69	1.41	1.81	0.95	0.83

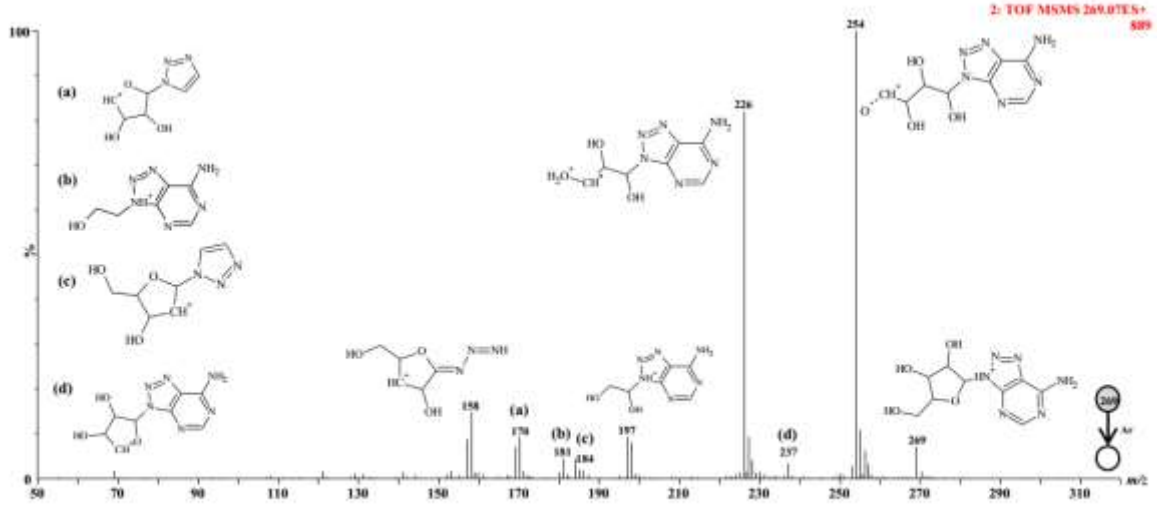
Fragmentation patterns of sugarcane metabolites identified in LC-ESI-MS/MS analysis using positive ionization. Fragmentation data was compared to Metlin database (https://metlin.scripps.edu/index.php) to find possible metabolites, and ACD/MS Structure ID suite software was used to compare fragmentation profiles to theoretically fragmented metabolites from database and manually checked.

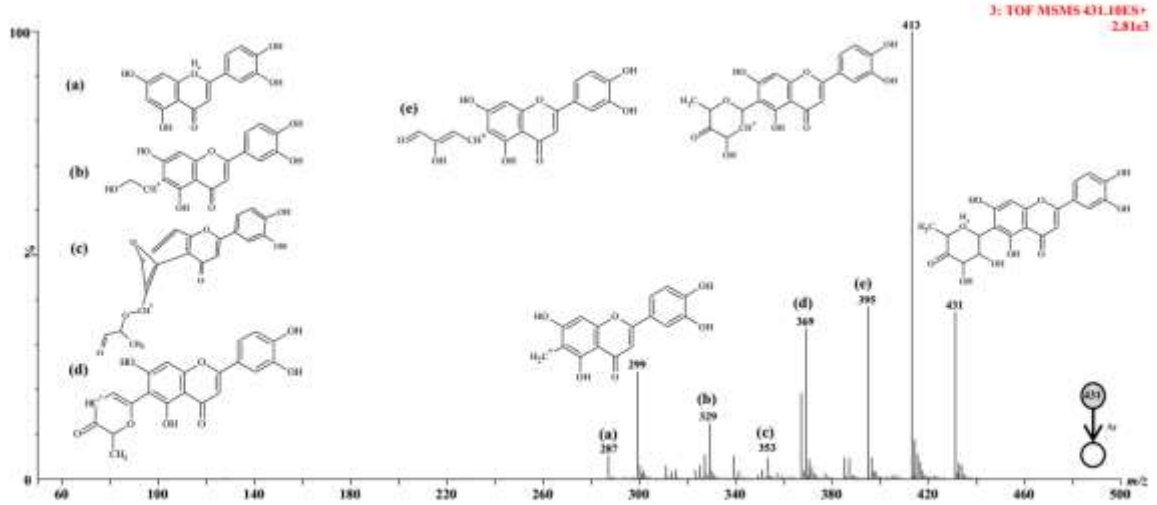


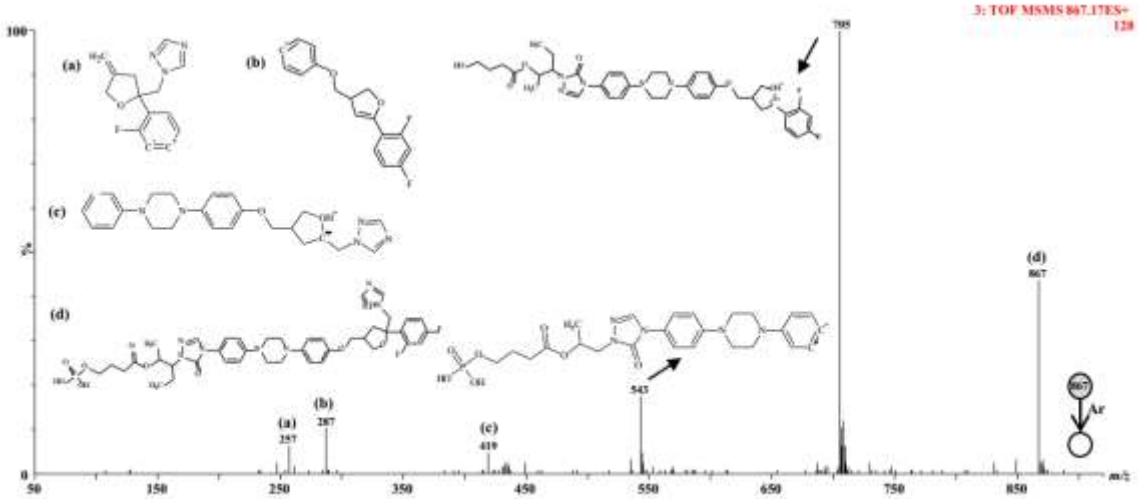


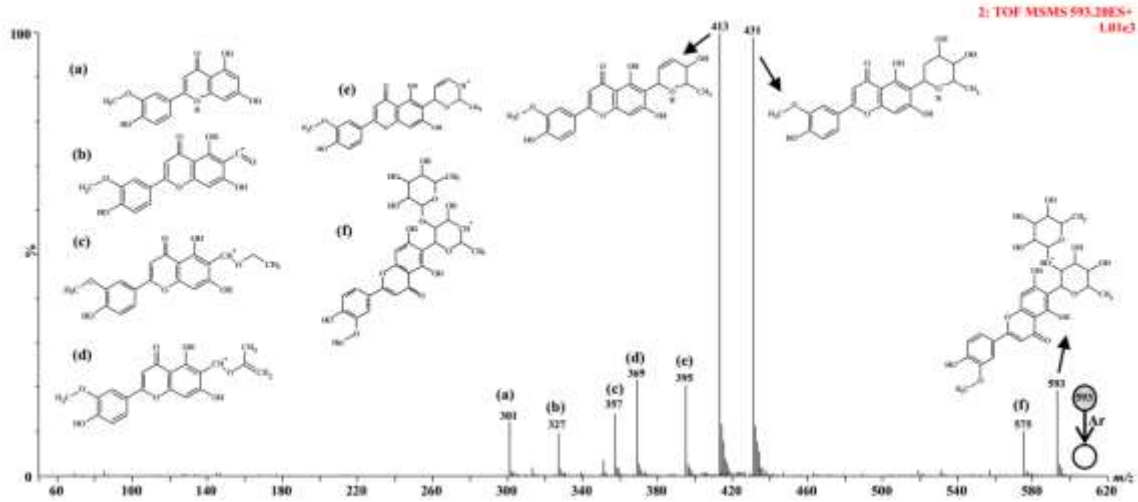


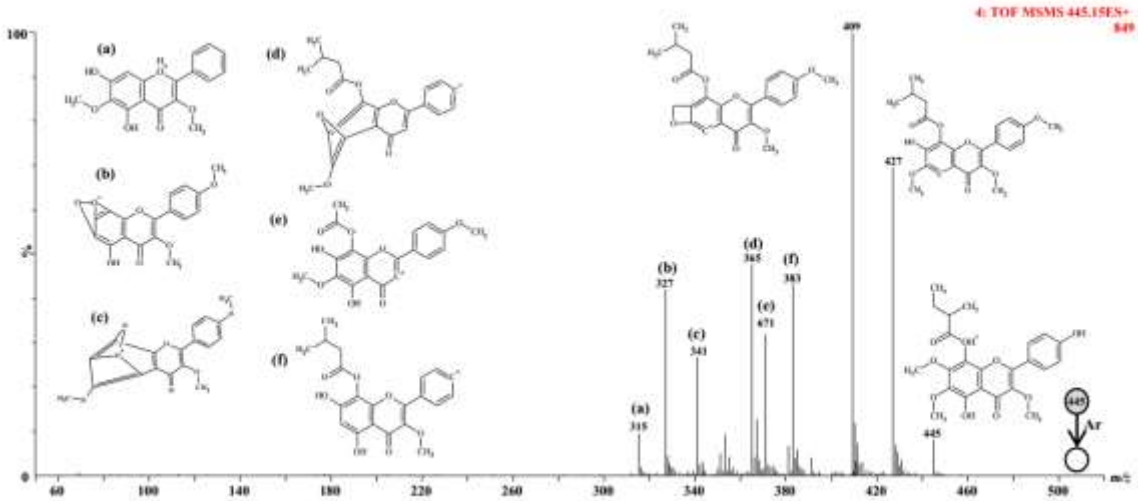

5: TOF MSMS 880.28ES+
8.30e3
100
%
0
50
150
250
350
450
550
650
750
850
m/z
557
177
289
845
880

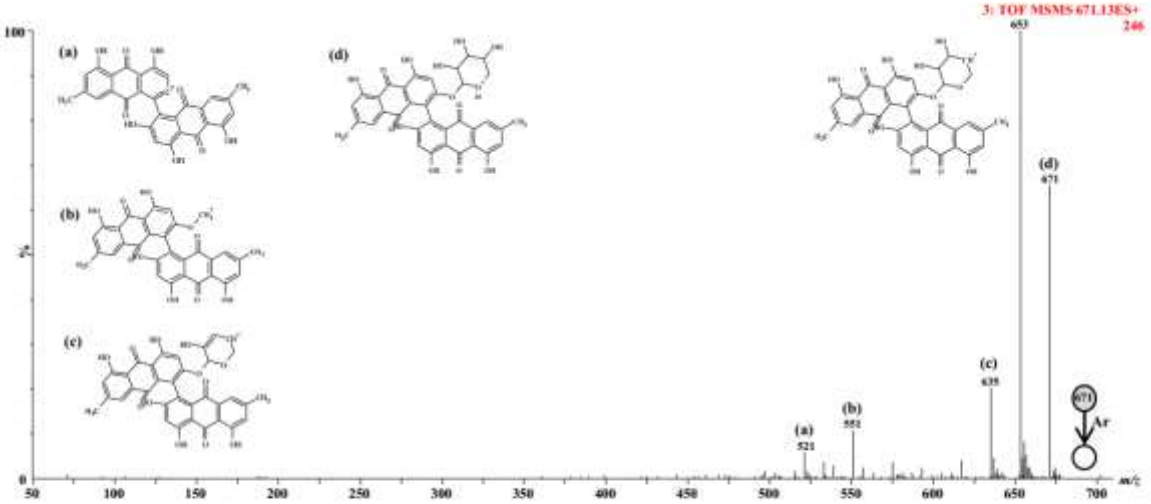

2: TOF MSMS 269.07ES+
889
100
%
0
50
70
90
110
130
150
170
190
210
230
250
270
290
310
m/z
(a)
(b)
(c)
(d)
158
170
181
184
197
226
237
254
269

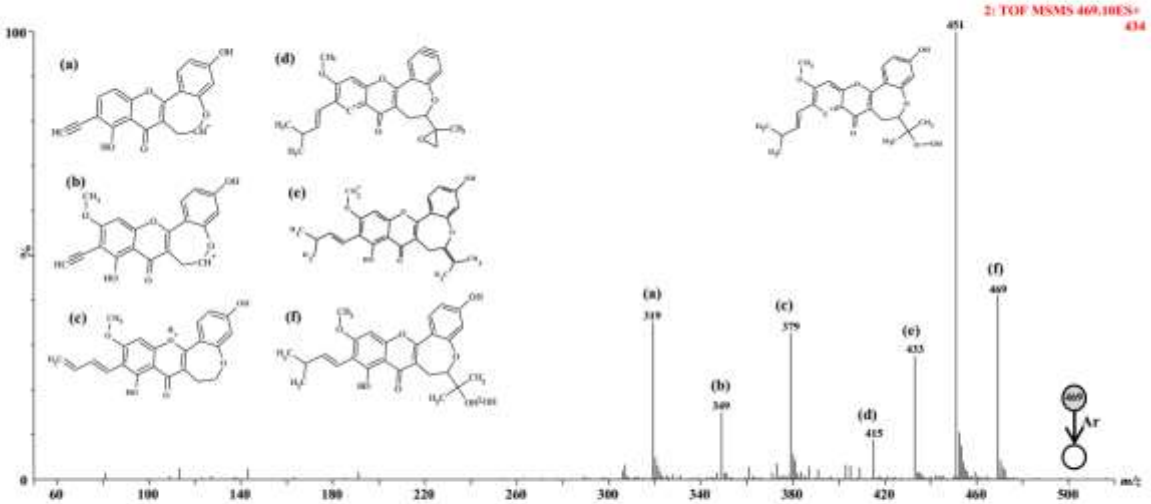

3: TOF MSMS 431.10ES+
2.81e3
100
%
0
60
100
140
180
220
260
300
340
380
420
460
500
m/z
(a)
(b)
(c)
(d)
(e)
287
299
329
353
369
395
413
431

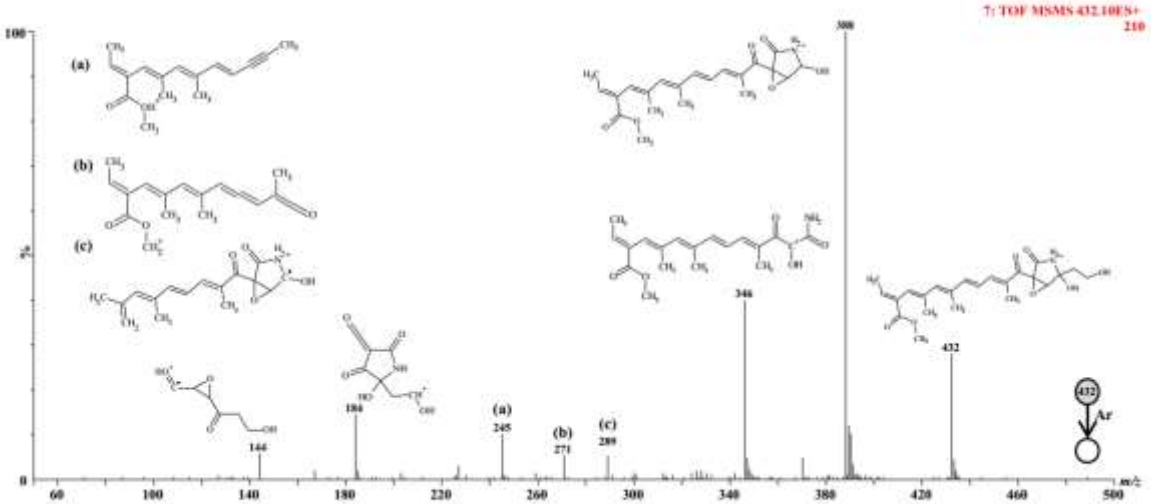

3: TOF MSMS 867.17ES+
128
100
%
0
(a)
(b)
(c)
(d)
257
287
419
543
705
867
867
Ar
50
150
250
350
450
550
650
750
850
m/z

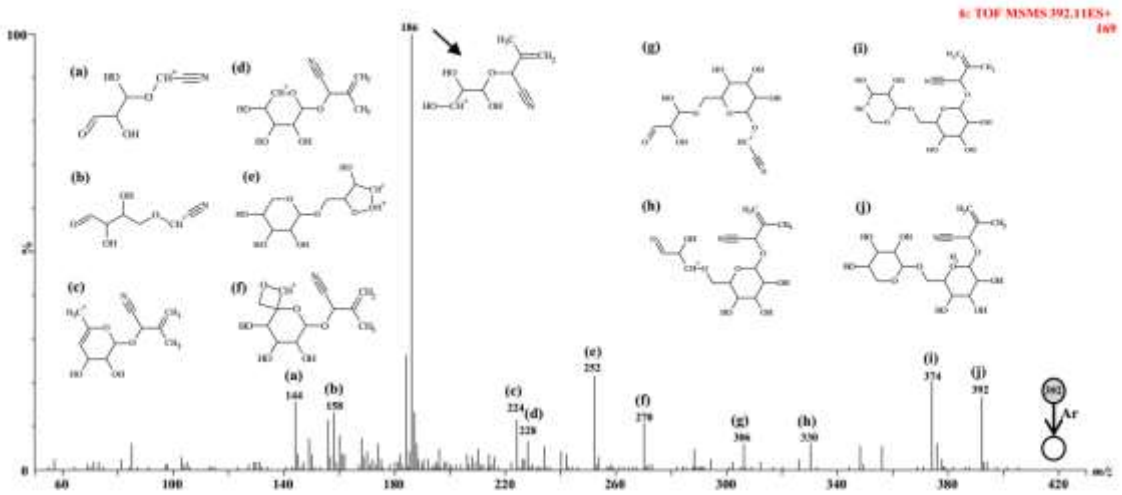

2: TOF MSMS 593.20ES+
1.01e3
100
%
0
(a)
(b)
(c)
(d)
(e)
(f)
301
327
357
369
395
413
431
593
593
Ar
60
100
140
180
220
260
300
340
380
420
460
500
540
580
620
m/z

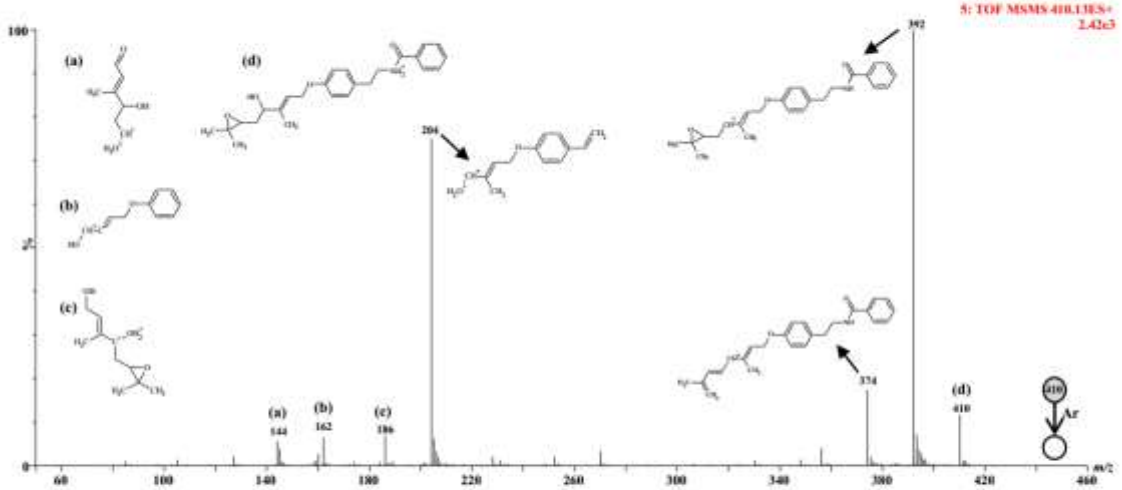

4: TOF MSMS 445.15ES+
849
100
%
0
(a)
(b)
(c)
(d)
(e)
(f)
315
327
341
345
371
383
409
427
445
445
Ar
60
100
140
180
220
260
300
340
380
420
460
500
m/z

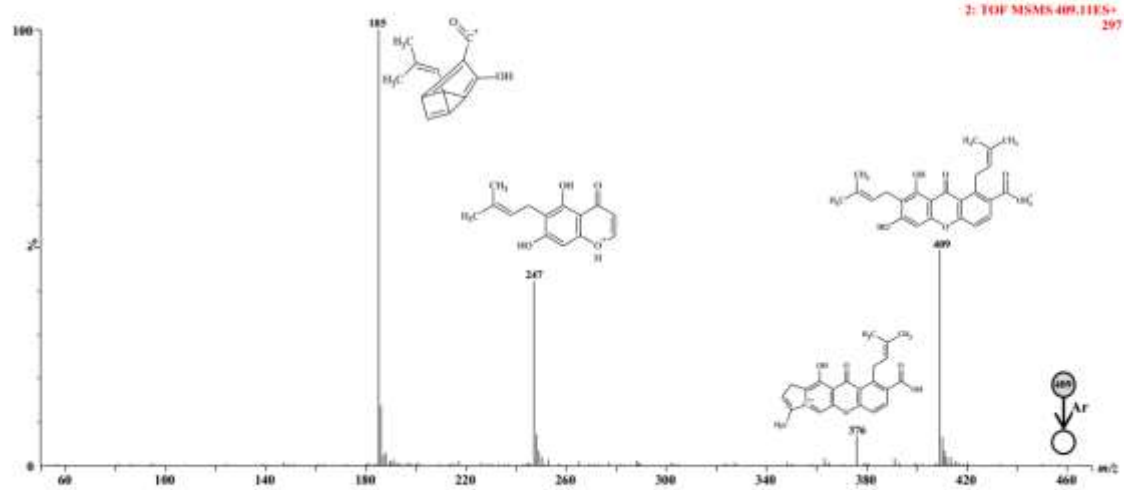

3: TOF MSMS 671.13ES+
653
246
(a)
521
(b)
551
(c)
635
(d)
671
671
Ar
m/z

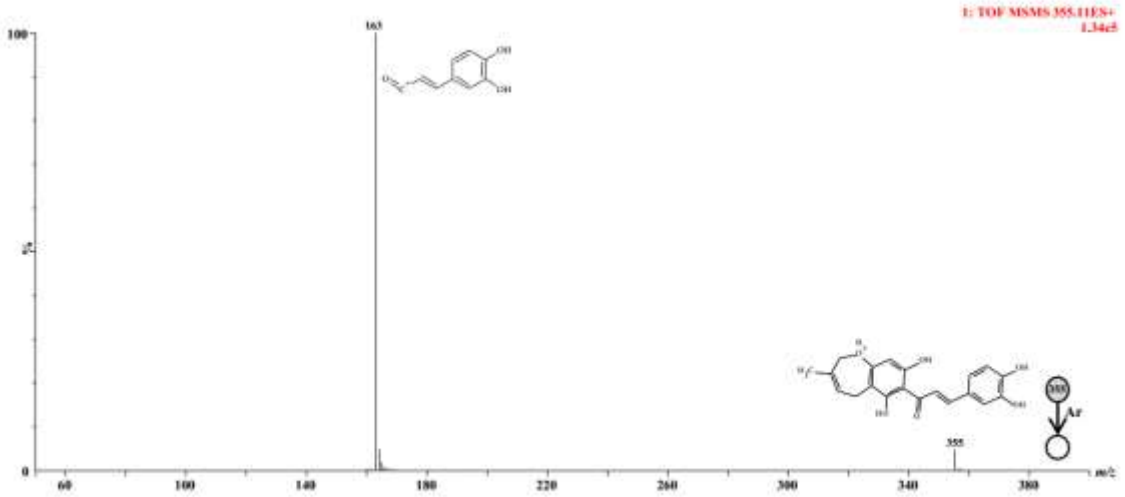

2: TOF MSMS 469.10ES+
434
451
(a)
319
(b)
349
(c)
379
(d)
415
(e)
433
(f)
469
469
Ar
m/z

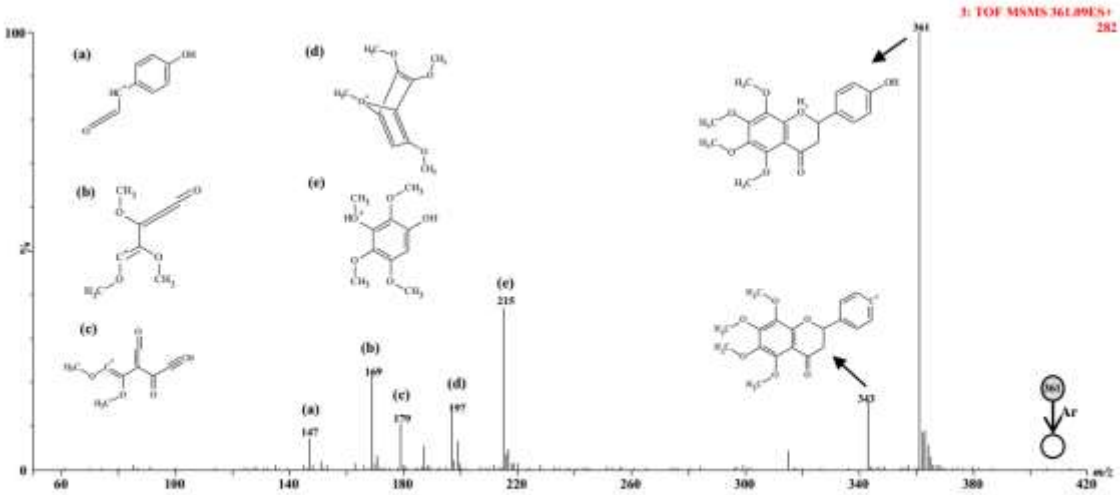

7: TOF MSMS 432.10ES+
210
384
346
432
432
Ar
(a)
245
(b)
271
(c)
289
184
146
m/z

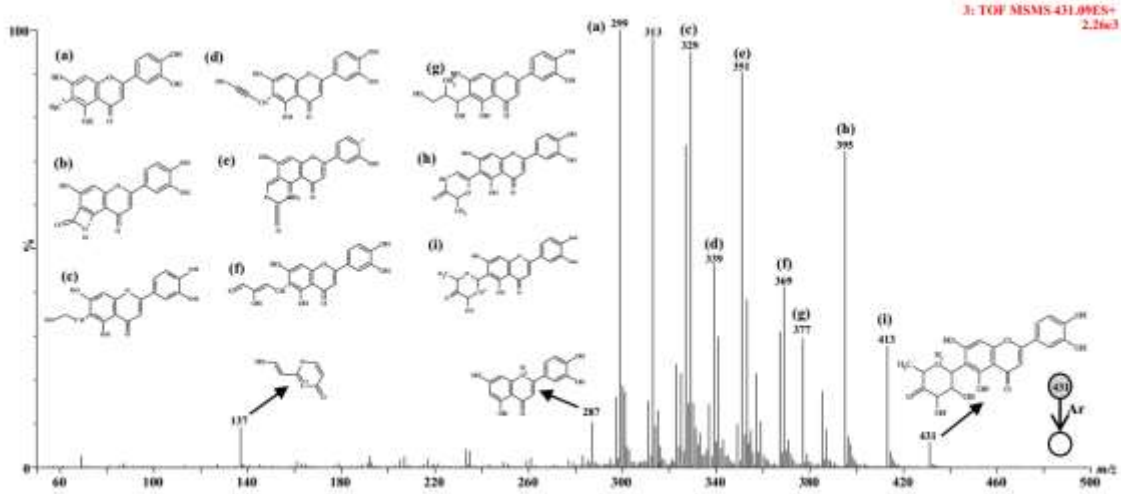

6: TOF MSMS 392.11ES+
169
100
%
0
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
186
144
158
224
228
252
270
306
330
374
392
Ar
60
100
140
180
220
260
300
340
380
420
m/z

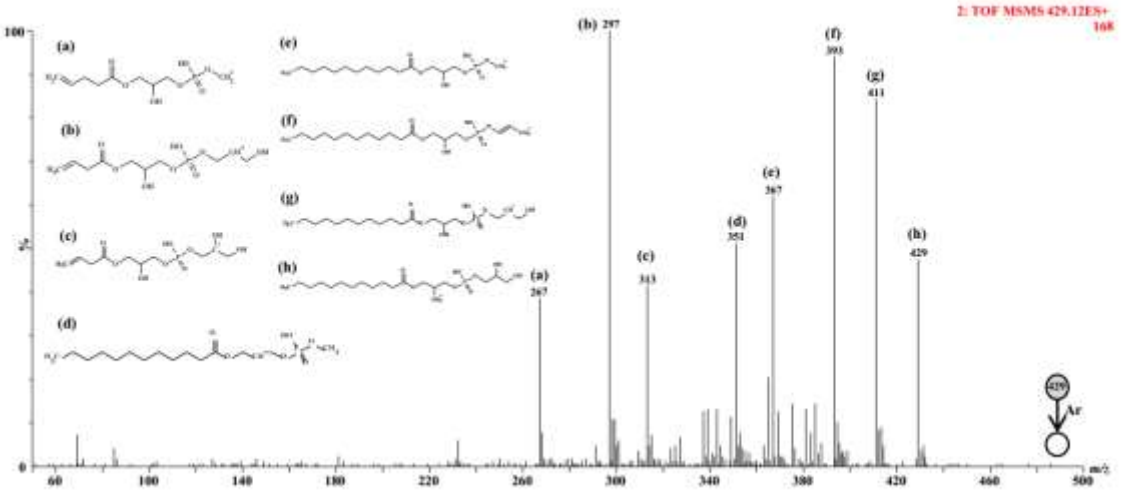

5: TOF MSMS 410.13ES+
2.42e3
100
%
0
(a)
(b)
(c)
(d)
392
204
374
410
144
162
186
Ar
60
100
140
180
220
260
300
340
380
420
460
m/z

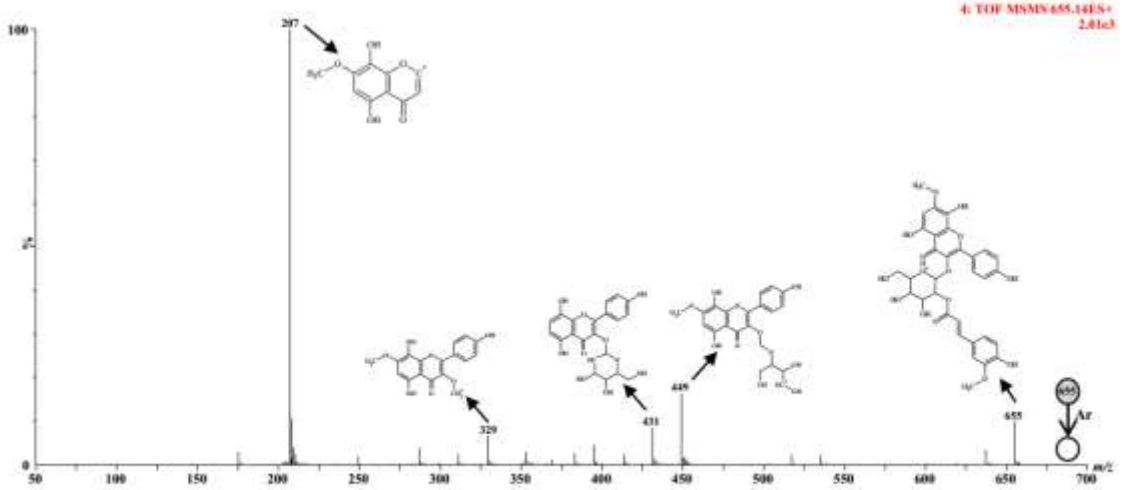

2: TOF MSMS 409.11ES+
297
100
%
0
185
247
376
409
Ar
60
100
140
180
220
260
300
340
380
420
460
m/z

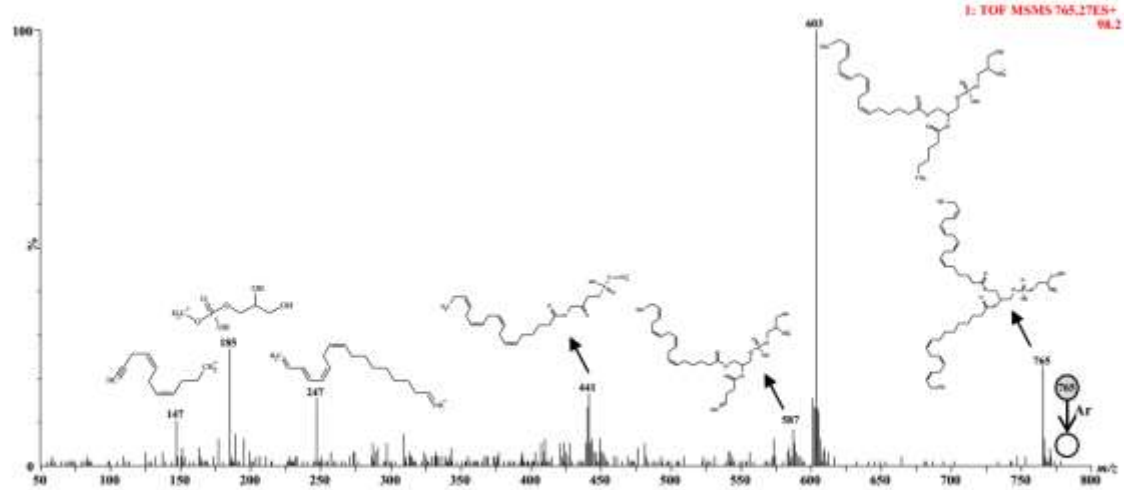

1: TOF MSMS 355.11ES+
1.34e5
100
%
0
163
355
Ar
60
100
140
180
220
260
300
340
380
m/z

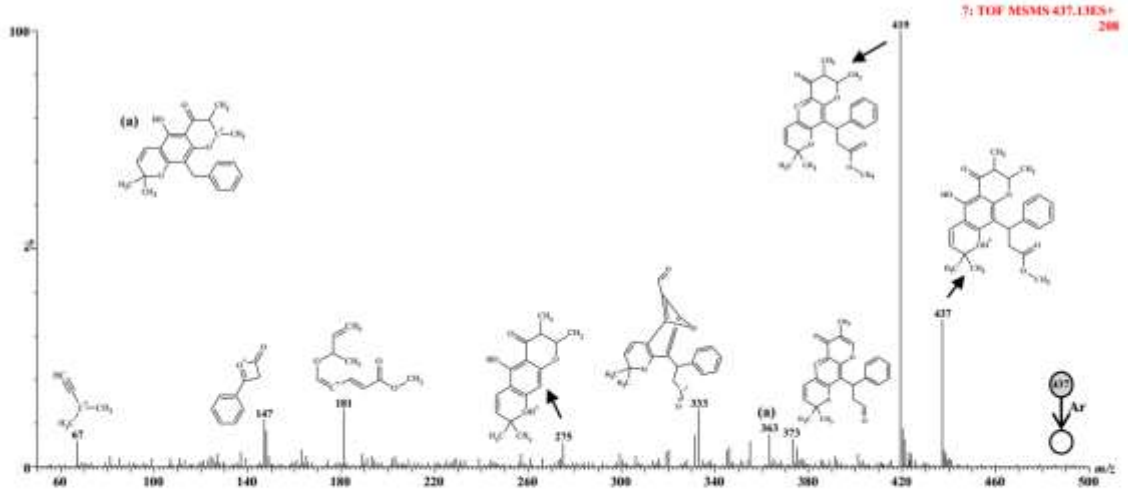

3: TOF MSMS 361.09ES+
282
100
%
0
(a)
(b)
(c)
(d)
(e)
147
169
179
197
215
343
361
Ar
60
100
140
180
220
260
300
340
380
420
m/z

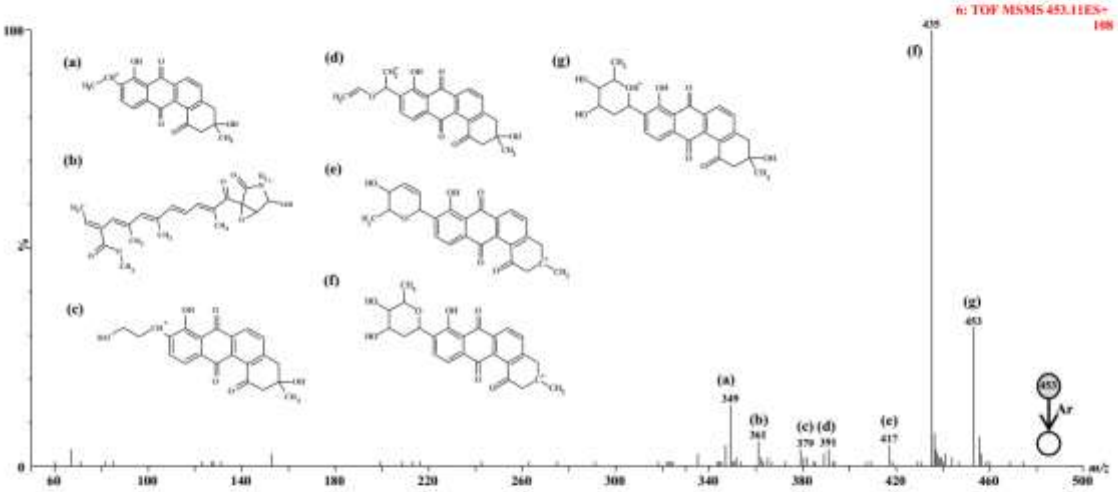

3: TOF MSMS 431.09ES+
2.26e3
100
%
0
(a) 299
313
(c) 328
(e) 351
(h) 395
(d) 359
(f) 369
(g) 377
(i) 413
137
287
431
Ar
60
100
140
180
220
260
300
340
380
420
460
500
m/z

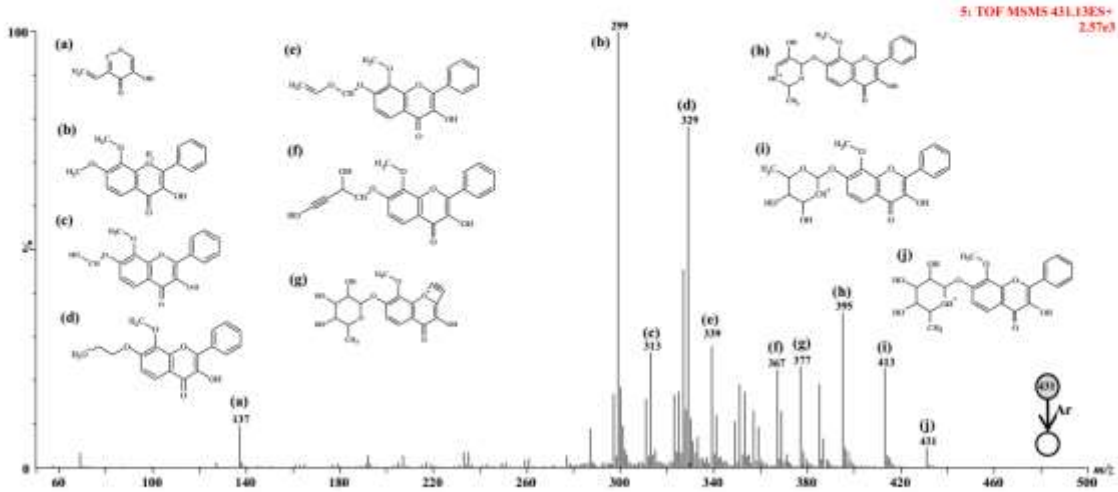

2: TOF MSMS 429.12ES+
168
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(b) 297
(c) 313
(d) 351
(e) 367
(f) 393
(g) 411
(h) 429
(a) 267
429
Ar
100
%
0
60
100
140
180
220
260
300
340
380
420
460
500
m/z

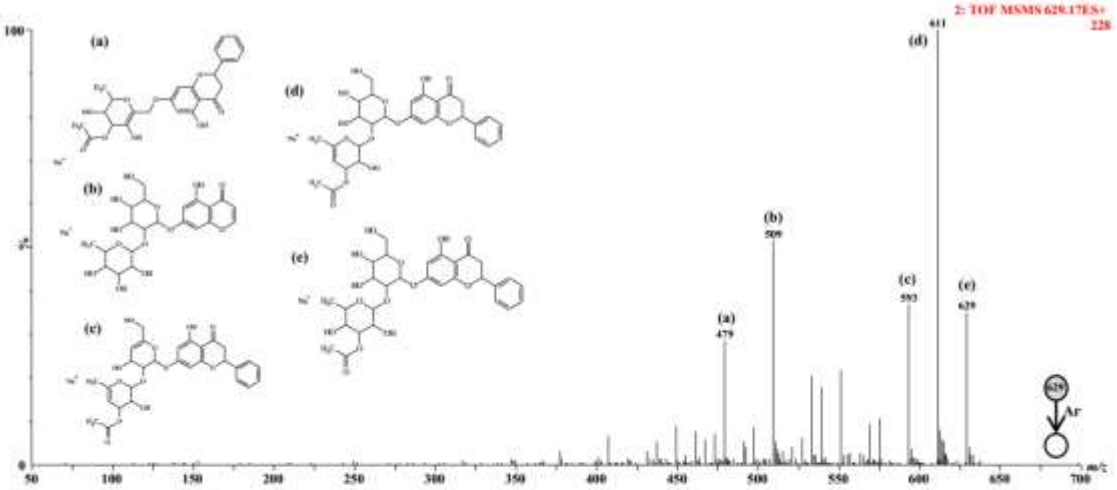

4: TOF MSMS 655.14ES+
2.01e3
297
329
431
449
655
655
Ar
100
%
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
m/z

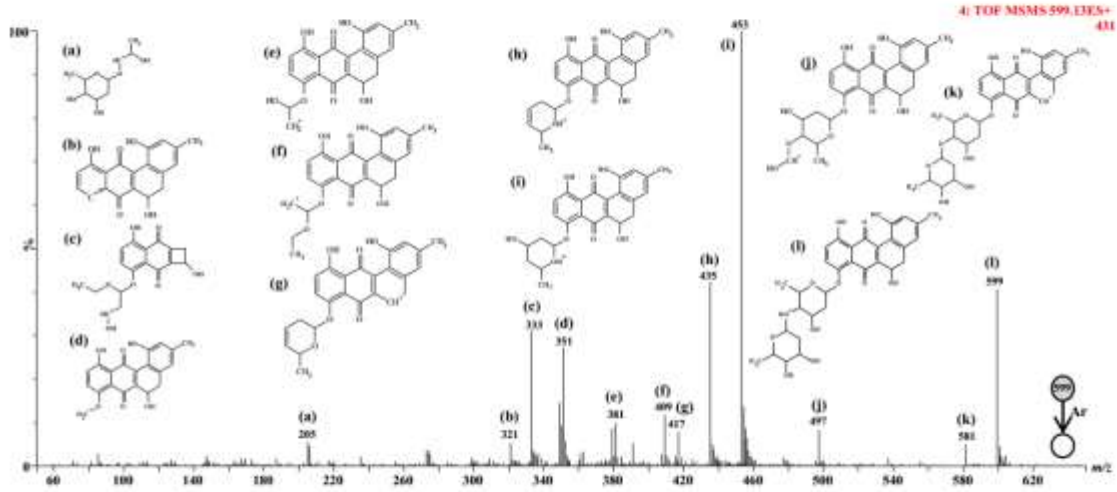

1: TOF MSMS 765.27ES+
98.2
603
185
147
247
441
587
765
765
Ar
100
%
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800
m/z

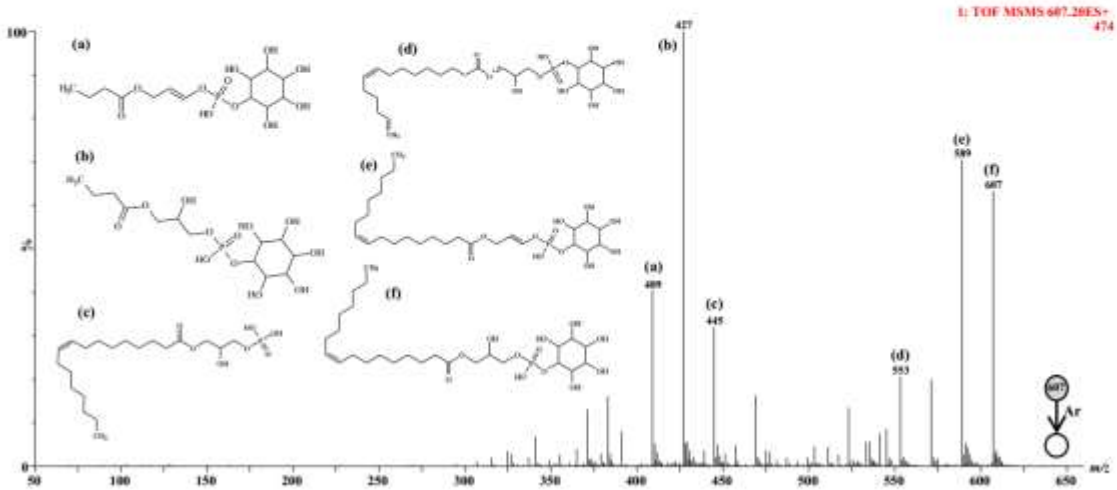

7: TOF MSMS 437.13ES+
208
100
%
0
(a)
67
147
181
275
335
(a)
363
373
419
437
437
Ar
60
100
140
180
220
260
300
340
380
420
460
500
m/z

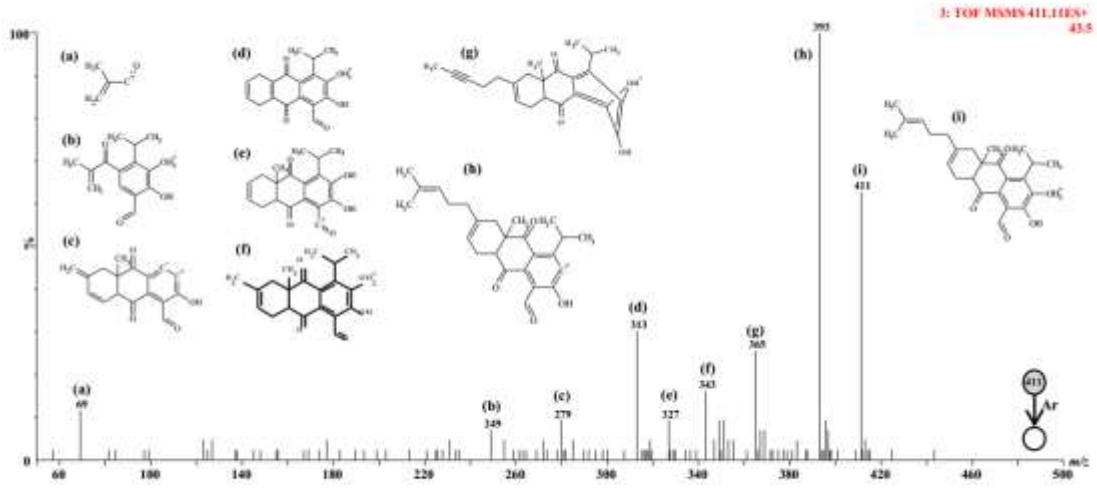

6: TOF MSMS 453.11ES+
108
100
%
0
(a)
(b)
(c)
(d)
(e)
(f)
(g)
435
(f)
(g)
453
(a)
349
(b)
361
(c) (d)
379 391
(e)
417
453
Ar
60
100
140
180
220
260
300
340
380
420
460
500
m/z


5: TOF MSMS 431.13ES+
2.57e3
100
%
0
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
299
(b)
(d)
329
(a)
137
(c)
313
(e)
339
(f)
367
(g)
377
(h)
395
(i)
413
(j)
431
431
Ar
60
100
140
180
220
260
300
340
380
420
460
500
m/z

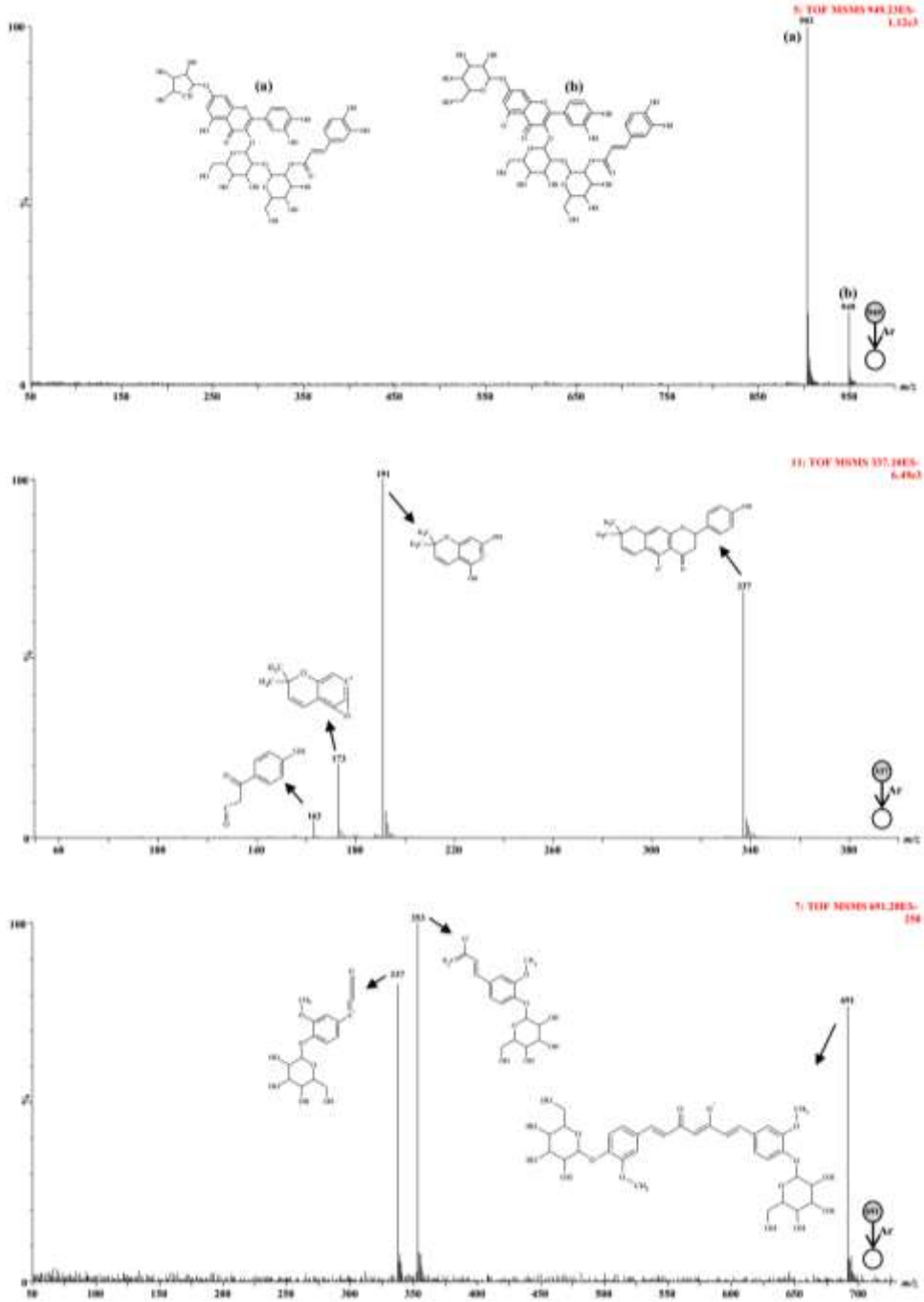

2: TOF MSMS 629.17ES+
228
(a)
(b)
(c)
(d)
(e)
479
509
592
611
629
Ar
m/z

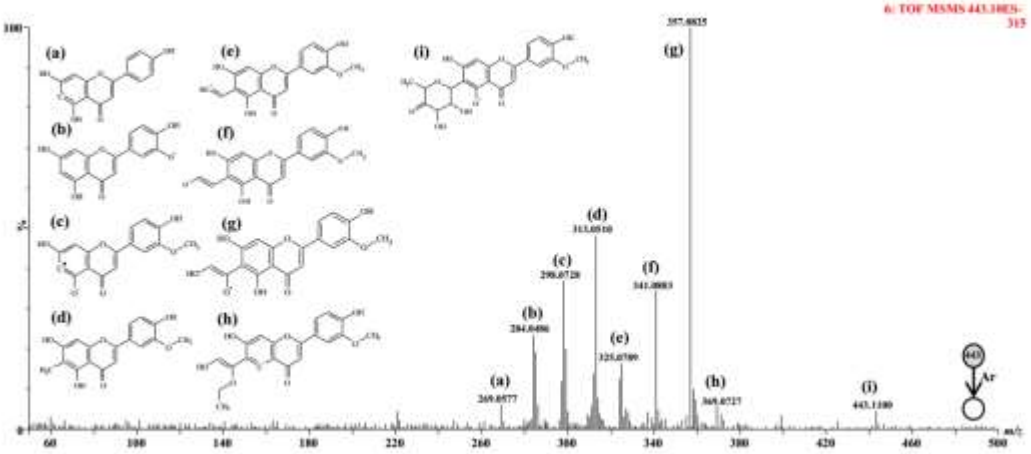

4: TOF MSMS 599.13ES+
431
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
205
321
333
351
381
399
417
435
453
497
581
599
Ar
m/z

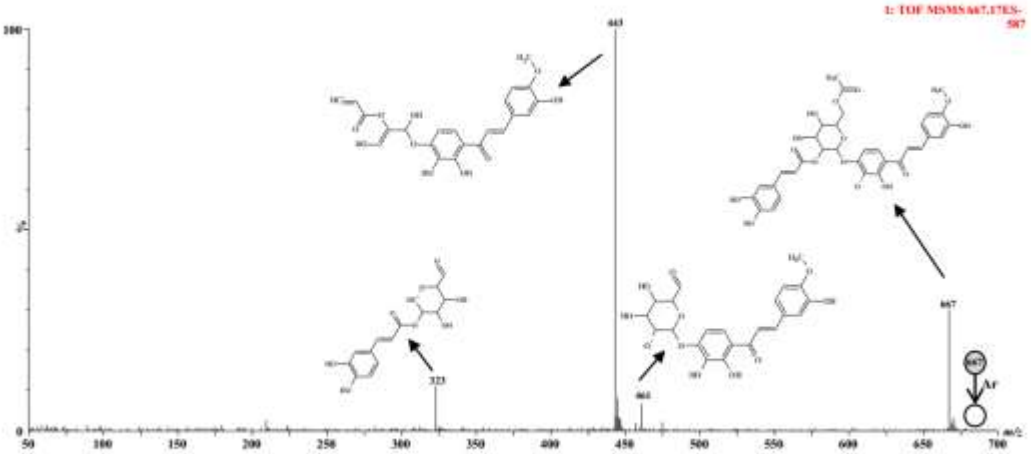

1: TOF MSMS 607.20ES+
474
(a)
(b)
(c)
(d)
(e)
(f)
409
427
445
553
589
607
Ar
m/z

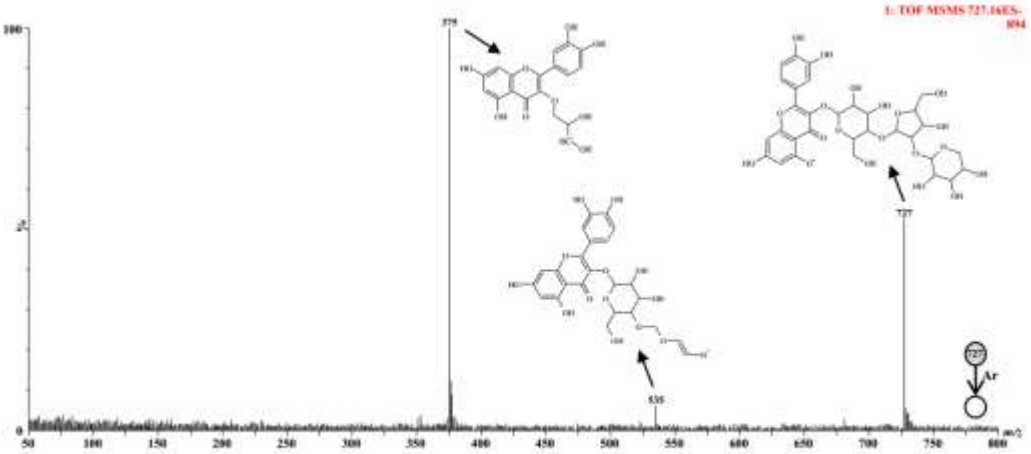

3: TOF MSMS 411.1 ES+
43.5
100
%
0
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
395
411
343
365
327
343
279
249
69
Ar
m/z
60
100
140
180
220
260
300
340
380
420
460
500

APPENDIX C. Fragmentation patterns of sugarcane metabolites identified in LC-ESI-MS/MS analysis using negative ionization. Fragmentation data was compared to Metlin database (https://metlin.scripps.edu/index.php) to find possible metabolites, and ACD/MS Structure ID suite software was used to compare fragmentation profiles to theoretically fragmented metabolites from database and manually checked.


(a)
(b)
(a)
(b)
191
173
147
337
353
337
691


6: TOF MSMS 443.10ES-
315
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
Ar


1: TOF MSMS 667.17ES-
587
443
323
667
Ar


1: TOF MSMS 727.16ES-
894
379
535
727
Ar