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

The complete genome sequence and comparative effectorome of *Sporisorium panici-leucophaei*: the causal smut disease agent in sourgrass (*Digitaria insularis*)

Gustavo Schiavone Crestana

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba
2020

Gustavo Schiavone Crestana
Biologist

The complete genome sequence and comparative effectorome of *Sporisorium panici-leucophaei*: the causal smut disease agent in sourgrass (*Digitaria insularis*)

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. PhD. **CLAUDIA BARROS MONTEIRO VITORELLO**

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba
2020

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

Crestana, Gustavo Schiavone

The complete genome sequence and comparative effectorome of *Sporisorium panici-leucophaei*: the causal smut disease agent in sourgrass (*Digitaria insularis*) / Gustavo Schiavone Crestana. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2020.

60 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Carvão do capim amargoso 2. Genômica comparativa 3. Biologia de efetores 4. Interação planta-patógeno I. Título

With all my love and affection,
To my dearest parents Silvio and Rita Helena
To my beloved sister Livia Helena

ACKNOWLEDGMENTS

This work was only possible through the collaboration and support of various people and institutions.

Firstly, I acknowledge and thank the University of São Paulo and specially "Luiz de Queiroz" College of Agriculture campus (ESALQ) for all the support over the years, from an undergraduate student to this moment, the conclusion of my dissertation.

I thank all the funding institutions that helped me along this journey. I thank the São Paulo Research Foundation (FAPESP) for the regular financial support and scholarships (Grant number: 2018/24028-5). I also thank the National Council for Scientific and Technological Development (CNPq) - Grant number: 144979/2018-7 and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 for previous scholarships.

I specially thank my advisor Dr. Claudia Barros Monteiro Vitorello. Over the years, you have trained me in the most diverse aspects: professional, personal and emotional. I cannot describe in words, all my affection for your figure as a genius scientist, strong woman and benevolent friend. Thank you for this wonderful journey.

To all my lab partners: Renato Bombardelli, Clesivan Pereira, Mariana Marrafon, Hugo Rody, Juliana Calderan, Deepak Sehgal, Laina Oliveira, Marcella Ferreira, Jéssica Mendes, Gustavo Lima, Jackeline Borba, Pedro Vilanova, Mariana Strabello, Juliana Benevenuto, Patricia Schaker, Natalia Teixeira, Andressa Bini, Leila Peters,Giselle Carvalho, Suzane Saito, Filipe Salveti, Natalia Moraes, Daniel Longatto, Sintia Almeida, I thank for their patience, friendship and for transforming a working environment into a family. Science is much better with you guys.

To all my university friends, Renato Bombardelli, Leonardo Piacenti, Vinícius Nicoletti, Pedro Beretta, Hugo Rody, Leandro Lemos, Pedro Catto, Felipe Lieber, for their sincere friendship and epical moments.

I thank all professors from de Department of Genetics, for their valuable lessons and for their contribution towards my professional training. A special thanks to Dr. Maria Carolina Quecine Verdi for her friendship.

I thank Dr. Juliana Benevenuto and Dr. Giselle Carvalho for initiating this project and for their support along this journey.

I thank the board of examiners for promptly accepting my invitation and for contributing to this work and to my professional training.

I thank the ESALQ Library staff and the Graduate Service (SVPG) for their help over this period.

I thank Dr. João Paulo Kitajima and Lucas Taniguti for their scientific support and incredible collaboration towards this project.

To Prof. Marie-Anne Van Sluys and Tatiana Corrêa, I thank for their assistance in Sanger sequencing.

I thank my family. And specially, I thank my parents Silvio Crestana and Rita Helena Schiavone, and my sister Livia Helena, for being my safe haven in all moments. Surely, I wouldn’t be able to accomplish this work without their emotional support. I also thank Mariana Lamy, my partner in this journey. Thank you for all your affection, encouragement and love.

“The best way to predict your future is to create it”

(Peter Drucker)

CONTENTS

RESUMO 7
ABSTRACT 8
PREFACE 9

CHAPTER 1: Sourgrass smut disease, an overview 11

1. Sourgrass (*Digitalia insularis*) 11
2. Sourgrass smut disease pathosystem 12
3. Effector biology 14

REFERENCES 16

CHAPTER 2: Smut pathogens: a comparative genomic approach, focusing on the causal agent of sougrass smut disease (*Sporisorium panici-leucophaei*) 23

1.Introduction 23
2. Material and Methods 26
2.1 Biological material 26
2.2 Interspecific hybridization assay for mating-type identification 27
2.3 Characterization of *S. panici-leucophaei* morphological structures 27
2.4 *In planta* infection time course 29
2.5 *S. panici-leucophaei* quantification (qPCR-assay) in *D. insularis* 29
2.6 Fungal DNA extraction 30
2.7 ITS-based phylogeny 30
2.8 Whole-genome sequencing strategies 30
2.9 Genome Assembly 31
2.10 Genome Annotation 32
2.11 Predicted Secretome and Effectorome 32
2.12 Orthologs group analysis 32
3. Results 34
3.1 Structures of *S. panici-leucophaei* 34
3.2 *S. panici-leucophaei* growth in planta 36
3.3 Interspecific Hybridization assay 38
3.4 ITS-based phylogeny 39
3.5 Genome Assembly 40
3.6 Genome and Protein Annotations 42
3.7 Comparative Genomics 44
3.8 Comparative Secretome and Effectorome 46
4. Discussion 47
4.1 Experimental approaches to study *S. panici-leucophaei* biology 47
4.2 An overview of *S. panici-leucophaei* genome features 48
4.3 Overview of the predicted protein *pool* of *S. panici-leucophaei* compared with *S. scitamineum* 49
4.4 Effectorome of *S. panici-leucophaei* 50
5. Conclusion 51

REFERENCES 51
SUPPLEMENTARY 58

RESUMO

A sequência completa do genoma e o efetoroma comparativo de *Sporisorium panici-leucophaei*: o agente causador do carvão do capim-amargoso (*Digitaria insularis*)

Fungos da família Ustilaginaceae são os principais agentes causais de doenças do carvão em plantas da família Poaceae, resultando em perdas agronômicas e econômicas significativas. As espécies causadoras de carvão apresentam estilo de vida biotrófico, dependente de uma íntima interação com o hospedeiro para completar seu ciclo de vida e produzir esporos (2n). O fungo *Sporisorium panici-leucophaei* (Spl) infecta e coloniza plântulas de *Digitaria insularis*, causando a doença denominada carvão do capim-amargoso. Apesar de possuir ciclo de vida semelhante à de outros carvões, poucas informações estão disponíveis sobre esse patossistema na literatura científica. O principal sintoma da doença causada por *S. panici-leucophaei* é mais semelhante ao carvão da cana-de-açúcar do que a de outras espécies de gramíneas. O desenvolvimento de uma estrutura em forma de chicote a partir do apex de plantas de capim-amargoso revelam potencial para comparar os patossistemas. Nesse sentido, o presente trabalho avaliou os padrões de colonização *in planta*, o ciclo de vida e formação de estruturas típicas penetração e colonização de *S. panici-leucophaei.* Além disso, o sequenciamento completo do genoma de *S. panici-leucophaei* foi concluído utilizando duas plataformas de sequenciamento distintas. A caracterização do ciclo de vida de *S. panici-leucophaei in vitro* foi concluída com a detecção das fases de germinação do esporo, formação do probasídio, formação de leveduras e geração da hifa dicariótica. Além disso, foi possível estabelecer quantitativamente o *time-course* de infecção *in planta* do patógeno através da germinação dos esporos 8 horas após a inoculação (hpi); formação do apressório às 48 hpi e detecção de hifas às 72 hpi. O genoma de *S. panici-leucophaei* foi concluído gerando uma montagem com 22 cromossomos mais o DNA mitocondrial. Alem disso, 6,404 genes foram preditos, dos quais, 382 foram secretados juntamente com 66 genes candidatos a efetores, padrão similar ao encontrado para o fungo *S. scitamineum*. As análises genômicas comparativas com o genoma do carvão da cana-de-açúcar geraram os primeiros insights sobre o conteúdo gênico compartilhado e específico da espécie, além de fornecer informações úteis sobre um patógeno de planta pouco estudado.

Palavras-chave: Carvão do capim-amargoso, Sequenciamento de nova geração, Genômica comparativa, Biologia de efetores, Interação planta-patógeno

ABSTRACT

The complete genome sequence and comparative effectorome of *Sporisorium panici-leucophaei*: the causal smut disease agent in sourgrass (*Digitaria insularis*)

Fungi of the Ustilaginaceae family are the main causal agents of smut disease in Poaceae plants, resulting in significant agronomic and economic losses. Smut species have a biotrophic lifestyle, which depends on a close interaction with a host to complete their life cycle and produce spores (2n). The fungus *Sporisorium panici-leucophaei* (Spl) infects and colonizes *Digitaria insularis* seedlings, causing the disease known as sourgrass smut disease. Despite presenting a similar life cycle compared to other smuts, scarce information is available about this pathosystem in the scientific literature. The development of a whip-like structure from the apex of sourgrass plants reveals potential for comparing with the sugarcane smut disease pathosystem. In this sense, the present work evaluated the colonization patterns *in planta*, the life cycle and formation of typical penetration structures of *S. panici-leucophaei*. In addition, complete genome sequencing of the *S. panici-leucophaei* was completed using two distinct sequencing platforms. The characterization of *S. panici-leucophaei* life cycle *in vitro* was concluded with the detection of distinct phases, such as spore germination, probasidium formation, sporidia formation and dikaryotic hyphae visualization In addition, it was possible to quantitatively establish the *in planta* infection *time-course* by spore germination 8 hours post-inoculation (hpi); appressorium formation at 48 hpi and hyphae detection at 72 hpi. The *S. panici-leucophaei* genome was completely assembled by generating a 22 chromosome assembly plus mitochondrial DNA. In addition, 6,404 genes were predicted, of which 382 were secreted along with 66 candidate effector genes, a pattern similar to that found for the *S. scitamineum* fungus. Comparative genomic analyses with the sugarcane smut genome yielded the first insights into shared and specific gene-content and provided useful information on a poorly studied plant pathogen.

Keywords: Sourgrass smut, Next generation sequencing (NGS), Comparative genomics, Effectors biology, Plant-pathogen interaction

PREFACE

This dissertation is a requirement to obtain the degree of Master in Agronomy (Genetics and Plant Breeding) at University of São Paulo (USP), campus — "Luiz de Queiroz" College of Agriculture (ESALQ) , Piracicaba, São Paulo, Brazil.

The research described above was conducted at the Department of Genetics under the supervision of Professor Claudia Barros Monteiro-Vitorello between March 2018 and December 2019. The dissertation is divided into two chapters which aim to generate biological information about the poorly explored pathosystem *Sporisorium panici-leucophaei vs Digitaria insularis*.

The first chapter consists of a bibliographic review on: the host plant *D. insularis*, known as sourgrass; the scarce information available about the complex interaction between the smut fungus *S. panici-leucophaei vs* sourgrass, and general aspects concerning the current state of effector biology for pathogens.

The second chapter, written as a scientific article, details the experimental part of the research project. Always in comparison with the sugarcane smut *Sporisorium scitamineum*, the chapter explores how both pathogens share similar life cycles and infection structures. In addition, the complete genome sequencing of *S. panici-leucophaei* was concluded in order to perform comparative genomic analyzes against *S. scitaminuem*, as a strategy to deepen the current knowledge about the exclusive and shared gene-content of the species.

CHAPTER 1: Sourgrass smut disease, an overview

1.) Sourgrass (*Digitaria insularis*)

Commonly called sourgrass, *Digitaria insularis* (L.) Fedde is a perennial plant species native to tropical and subtropical regions of America (Takano *et al.*, 2018). This perennial weed, classified within the Poaceae family, presents a C4 photosynthetic metabolism (Kissmann, 1997). It is characterized by its medium port size (50 to 100 cm in height) with striped culms, long internodes, long leaves presenting pilose sheath, and membranous ligule (Lorenzi, 2000). *D. insularis* is a weed with aggressive and invasive behavior, presenting slow initial growth, particularly under shading conditions or low temperatures (Machado *et al.*, 2006; Machado *et al.*, 2008). However, its rapid adaptation allows increased survival to adverse environmental conditions, limiting the growth of other plant species (Brighenti & Oliveira, 2011). Rapid plant growth, coupled with high yield seed production and high germination rates, provide sourgrass specific features that classify it as a highly competitive species with enormous infesting weed potential (Machado *et al.*, 2006). In addition, the species reproduces by seeds and rhizomes, forming complex clumps from them (Kissmann, 1997).

In Brazil, sourgrass has been reported by Brandão *et al.* (1995) in various regions of the country, including the Cerrado biome, at both grains producing and pasture areas (Carvalho & Pitelli, 1992). Sourgrass has also been identified in banana and sugarcane crops in the Rio de Janeiro State (Oliveira & Freitas, 2008), as well as farming lands in São Paulo State (Kuva *et al.*, 2008). Gemelli *et al*. (2013) quantified the agronomic impact of sourgrass on “safrinha” corn crop in the municipality of Maringá, Paraná State, leading to a 35% decrease in yield production.

The management of *D. insularis* requires chemical control implemented up to 35 days post-emergence, before rhizome formation (Machado *et al.*, 2006; Correia & Durigan, 2009). Among agrochemicals used to control weeds, Glyphosate (n-phosphonomethyl glycine) is one of the most commonly used in Brazil and worldwide (Duke & Powles, 2008). Glyphosate acts systemically quick trough translocation from leaves to roots, rhizomes, and apical meristems (Franz, 1985; Gruys & Sikorslci, 1998). Its mechanism of action consists of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPs) inhibition, a catalyst enzyme responsible for shikimic acid and pyruvate phosphate condensation, thus preventing the synthesis of tyrosine, tryptophan, and phenylalanine, considered as essential amino acids (Jaworski, 1972;

Zablotowicz & Reddy, 2004). After Glyphosate application, the main symptoms consist of leaf chlorosis followed by necrosis, in addition to wrinkling and meristem and rhizomes necrosis. Despite the high efficacy of Glyphosate, sourgrass has gained relevance in recent years due to its high tolerance to the pesticide. First reports of resistance were observed in Paraguay by Timossi, Durigon and Leite (2006) and Machado *et al.* (2006), while in Brazil, the first report occurred in the municipality of Guaíra (Paraná State) on a soybean crop (Duke & Powels, 2008). The main costs of resistance are related to the need for alternative herbicide use and productivity losses due to competition from remaining resistant weeds in crop areas. Therefore, an increase in the total production costs in these agricultural areas is significant (Embrapa, 2017). As estimated by Embrapa - Technical Circular (2017), controlling Glyphosate-resistant sourgrass increases by 165% the production cost, resulting in an average a R$ 170.50/ha increase in crops such as soybean, for example.

2.) Sourgrass smut disease pathosystem

Despite its high aggressiveness and rapid growth, sourgrass is also a host to several pests and pathogens such as the maize chlorotic mottle and dwarf mosaic viruses, the nematode *Radopholus similis* and the moth *Mocis latipes*, the fungus *Sporisorium panici-leucophaei* and the bacteria *Xanthomonas citri* (Invasive Species Compendium, 2018). The basidiomycete S*porisorium panici-leucophaei* belongs to the Ustilaginaceae family which is commonly referred as smut fungus and comprises more than 600 species (Benevenuto *et al*., 2016). Formerly called *Lundquistia panici-leucophaei*, the pathogen was taxonomically reclassified by Cunnington *et al*. (2005) and Stoll, Begerow and Oberwinkler (2005) as the genus *Sporisorium* through phylogenetic and molecular analysis.

The hosts of *S. panici-leucophaei* belong to the genera *Digitaria*, *Echinochloa*, *Oplismenopsis,* and *Panicum*. However, *Digitaria insularis* is the primary host (Clayton & Renvoize, 1986; Piątek, 2006). Infected *D. insularis* plants were reported by Pérez *et al*. (2002) on all the American continent extension (North, Central and South). In Brazil, sourgrass smut was identified in the State of São Paulo in 1989 by Figueiredo *et al.* (1989). The life-cycle information of the pathogen and infection reports of *D. insularis* are scarce in the scientific literature. Nevertheless, many smut pathogens share similar life cycles (Sundar *et al*., 2012), thus suggesting a similarity to the *Sporisorium scitamineum* life cycle, a basidiomycete commonly known as sugarcane smut. *S. scitamineum* life cycle is well described (Sundar *et al*., 2012; Taniguti *et al*., 2015; Benevenuto *et al*., 2016). It begins when diploid teliospores (2n) germinate over the surface of sugarcane buds (internodes) (Fig. 1). Germination leads to

probasidium formation where meiosis produces four haploid (n) cells (or basidiospores), which present yeast-like growth *in vitro*. Sporidia of opposite sexual types (A and B) emit conjugation (reproductive) hyphae, thus enabling the occurrence of a sexual reaction leading to the anastomosis. The sexual reaction between compatible sporidia occurs through a pheromone and membrane receptors recognition system controlled by two loci, *a* and *b*; a process also described for other smut species, such as *Ustilago maydis*, *U. hordei* and *S. reilianum* (Bakkeren *et al*., 2008; Vollmeister *et al*., 2012). Dikaryotic hyphae then form the appressorium, a structure designed to penetrate the host plant tissues (Peters *et al*., 2017). *S. scitamineum* invades the host between 6 and 36 hours after teliospore germination (Sundar *et al*., 2012; Monteiro-Vitorello *et al*., 2018). After penetration of the stem buds and colonization of the apical meristem, *S. scitamineum* completes its life cycle, giving rise to new teliospores (2n). These spores are visible at the apex of the plant formed in a black appendix: a whip-like structure commonly described as the characteristic symptom of sugarcane smut disease (Sundar *et al*., 2012).

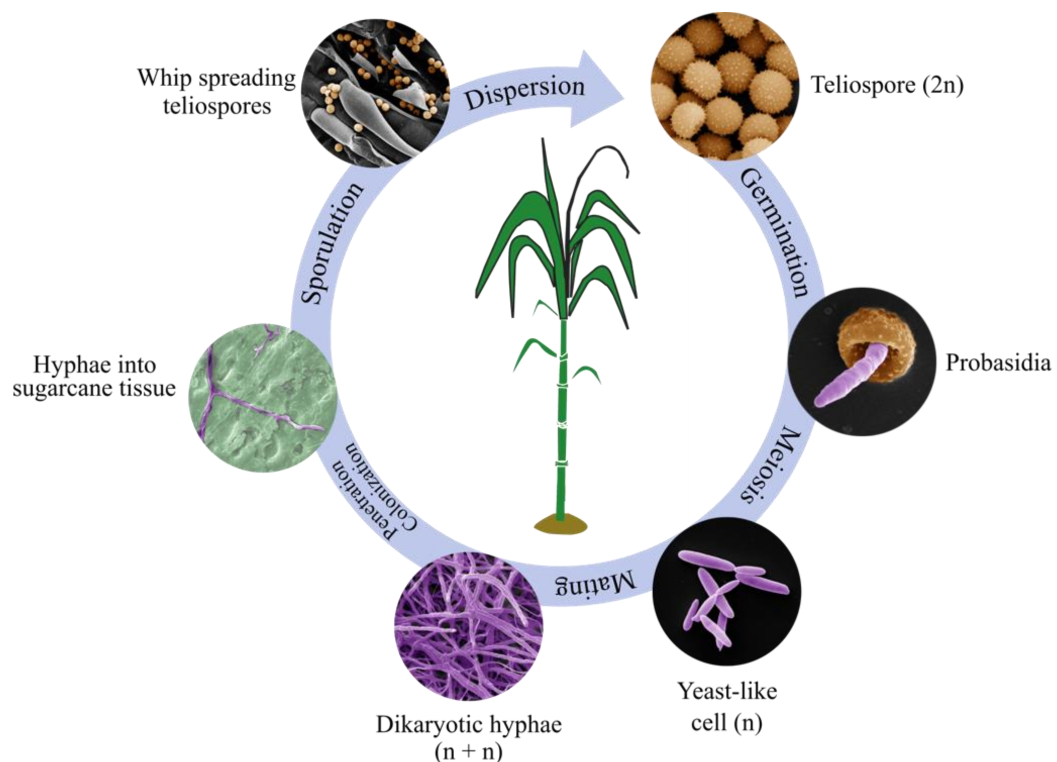

Figure 1. *Sporisorium scitamineum* in various developmental stages and within host tissues. Retrieved from Monteiro-Vitorello *et al*., 2018.

The symptom of sourgrass smut disease is very similar to that of sugarcane smut and different from other smut diseases. The whip-like structure developed from the apex of the plant produces the teliospores in the life cycle of S. *panici-leucophaei* and *S. scitamineum.*

Previous experiments suggest high similarity between the life cycles of sugarcane and sourgrass smuts. The accessibility of genome sequencing technologies has allowed large-scale analyses of an increasing number of plant-pathogen species. Comparative genomics is a powerful tool for deciphering and providing *insights* into conserved and specific pathogenicity mechanisms from distinct pathosystems (Plissonneau *et al*., 2017). Benevenuto *et al*. (2018) performed a comparative genomics analysis among nine smut species of different hosts, identifying lineage-specific genes, of which effector genes play an essential role in pathogen-host interaction. Effectors are molecules secreted by pathogens that modulate host physiological processes, acting directly or indirectly on the host immune system to allow colonization (Kamoun, 2006; Dalio *et al*., 2014; Selin *et al*., 2016). Despite the biological relevance of these molecules, there are no reports on the identification of candidate effector genes for the *S. panici-leucophaei vs* sourgrass interaction, thus impairing molecular understanding of which molecules/mechanisms play a relevant role in this pathosystem.

3.) Effector biology

Several filamentous fungi attack crops of great agronomic importance, generating economic losses by reducing productivity. Biotrophic fungi are organisms that overcome host defense responses without damaging the integrity of physiological processes, essential to the plant survival, as they require metabolically active plant tissues for their nutrition. In contrast, plants have developed defense mechanisms that recognize the invasion of pathogens (Dodds & Rathjen, 2010). Both plant defense and host attack systems are under constant and accelerated evolutionary pressure known as "*arms race*" (Giraldo & Valent, 2013). Consequently, the defense mechanisms encountered in both organisms have become highly complex and diverse.

Plants have developed strategies that enable the perception of pathogen invasion, of which two can be highlighted: *PAMP triggered immunity* (PTI) and *Effector triggered immunity* (ETI) (Jones & Dangl, 2006). PTI is related to the plant basal defense response through the recognition of specific molecules called Pathogen Associated Molecular Patterns (PAMPs) or Microbe Associated Molecular Patterns (MAMPs). These molecules are essential for microbial survival and are highly conserved among species (Medzhitov & Janeway, 1997). Often, such molecules originate from non-pathogenic microorganisms, so the term MAMP is also commonly used. PAMPs/MAMPs act as elicitor molecules and trigger innate defense responses. Among these molecules perceived by the plant, essential pathogens components are recognized, such as flagellin, found in the bacterial flagella and chitin and glucan, structural components found in the cell wall of fungi and oomycetes, respectively (Zipfel and Felix,

2005). These molecules are perceived in the host extracellular environment by molecular pattern recognition receptors (PRRs). There are two main types of transmembrane receptors: receptor-like proteins (RLPs) and receptor-like kinases (RLKs).

Both classes of proteins usually have leucine-rich repeats (LRRs) or a lysine motif (LysM) in the extracellular domain. By recognizing PAMPs/MAMPs, PRRs, activate plant PTI triggering responses as an attempt to combat microbial infection. Among these defense responses, calcium (Ca2+) influx into cells, the production of reactive oxygen species (ROS), gene expression involved in the biosynthesis of peptides and antimicrobial compounds, and callose deposition on the cell wall are important primary ones among hosts (Giraldo & Valent, 2013).

In order to override the PTI response triggered by the plant, pathogenic microorganisms developed effectors, and the plants, in turn, developed *Effector-triggered immunity* (ETI). First, effectors are secreted molecules that modulate physiological processes of the host. The most accepted definition for effectors involves duality in their function, depending on the host. They may be related to both pathogenicity and the activation of host resistance (Hogenhout *et al*., 2009; Dalio *et al*., 2014). Besides, these molecules are not necessarily proteins (Win *et al*., 2012). Weiberg *et al*. (2013) reported that microRNAs (sRNA) derived from *B. cinerea* could act as effectors by sabotaging the *Arabidopsis* interference RNA system. Effectors can operate in various ways, for example, by the inhibition of enzymes (hydrolases) in the apoplast, inhibition of PRR activation, inhibition of kinase-mediated signaling cascades, transcriptome reprogramming related to defense responses, degradation of defense-related components/proteins, among others (Dalio *et al*., 2014).

The perception of effectors triggers ETI and their subsequent suppression in the apoplast (apoplastic effectors) or in the host cell cytoplasm (cytoplasmatic effectors) (Hogenhout *et al*., 2009). Effectors are recognized by proteins encoded by the host resistance (R) genes. These receptors have NB-LRR (nucleotide binding-leucine rich repeat) domains responsible for effector binding and recognition/perception (Glazebrook, 2005). Effector variability may influence their recognition by the host. When enabled, the ETI triggers a number of responses from which the main ones are hypersensitive response (HR), transcriptional reprogramming, reactive oxygen species (ROS) production, activation of transcription factors and R genes, kinase signaling cascades, among others (Tsuda & Katagiri, 2010; Dalio *et al*., 2014).

Pathogens have developed strategies to avoid recognition by resistance proteins with a constant selection of new effectors or modification of old ones. This coordinated interaction led

plants and pathogens to a complex process of coevolution, highly dynamic and variable, generating extreme diversification of resistance receptors and effectors (Dodds & Rathjen, 2010). The mechanisms described above are part of the Zigzag model proposed by Jones and Dangl (2006) that aimed to quantitatively characterize the interaction between plants and pathogens and describe which molecules are involved in the process. Although widely accepted by the scientific community in the conceptualization and elucidation aspects related to host-pathogen interactions, the Zigzag model has limitations.

To predict proteins potentially secreted by the *S. scitamineum* proteome, Taniguti *et al*. (2015) considered the presence of signal peptide and the absence of transmembrane domains and glycosylphosphatidylinositol (GPI) in the proteome. Based on these predictions, the secretome of the fungus comprised 305 proteins, including 54 CAZymes and 70 effector candidates. Barnabas *et al*. (2017), using computational tools, identified orthologous candidate effectors genes in *S. scitamineum* that were functionally characterized in *U. maydis* and made comparisons with other smut species. The results revealed a set of candidate effector genes conserved among smuts composing the smut core effectors (Uhse & Djamei, 2018). Despite the constant development of new technologies and models, several molecular mechanisms involved in plant *vs.* pathogen interactions are still poorly understood, especially for filamentous fungi (Sharpee & Dean, 2016). Effector biology is considered as an important tool in understanding the biology of biointeractions with a potential to influence breeding and plant protection programs with promising molecular-based selection strategies (Hogenhout *et al*., 2009; Petre *et al*., 2015; Sharpee & Dean, 2016). Due to the highly variable nature of effectors both *in silico* and function, case-by-case studies are essential to characterize these molecules in specific pathosystems. Here, it is described for the first time the complete genome sequence of *S. panici-leucophaei* and a comparative genomics approach to highlighting similarities to *S. scitamineum* biology.

REFERENCES

BAKKEREN G., KÄMPER J., SCHIRAWSKI J. (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. Fungal Genetics and Biology,v. 45,p. S15–S21.

BARNABAS L., ASHWIN N.M.R., RAMESH SUNDAR A., MALATHI P., VISWANATHAN R. (2017). Putative orthologs of *Ustilago maydis* effectors screened from the genome of sugarcane smut fungus - *Sporisorium scitamineum*. Australasian Plant Pathology, v. 46, p. 147-156.

BEGEROW D., GÖKER M., LUTZ M., STOLL M. (2004). On the evolution of smut fungi on their hosts. Frontiers in basidiomycote mycology, p. 81–98.

BEGEROW D., SCHÄFFER A.M., KELLNER R., YURKOV A., KEMLER M., OBERWINKLER F. (2014). Ustilaginomycotina. In: MCLAUGHLIN, D.; SPATAFORA, J. W. (Eds.).The Mycota. vol. 7. Part A. Systematics and evolution. 2. ed. Berlin: Springer, p. 295– 329.

BENEVENUTO J., LONGATTO D.P., REIS G.V., MIELNICHUK N., PALHARES A.C., CARVALHO G., *et al.* (2016). Molecular variability and genetic relationship among Brazilian strains of the sugarcane smut fungus. FEMS Microbiology Letters, v. 363, n. 24, p. fnw277.

BENEVENUTO J., TEIXEIRA-SILVA N.S., KURAMAE E.E., CROLL D., MONTEIRO-VITORELLO C.B. (2018). Comparative Genomics of Smut Pathogens: Insights From Orphans and Positively Selected Genes Into Host Specialization. Frontiers in Microbiology, v. 9, p. 660.

BRANDÃO M., PALUMA E., KEIN V.L.G., MAUTONE L., GUIMARÃES E.F., PEREIRA R.C., *et al.* (1995). Plantas daninhas do estado do Rio de Janeiro: acréscimo aos trabalhos já efetuados no estado. Planta Daninha, v. 13, n. 2, p. 98–116.

BRIGHENTI A. M., OLIVEIRA M. F. (2011). Biologia de Plantas Daninhas. [s.l.] Embrapa Milho e Sorgo-Capítulo em livro técnico-científico (ALICE).

CARVALHO S.L., PITELLI R.A. (1992). Levantamento e análise fitossociológica das principais espécies de plantas daninhas de pastagens da região de Selvíria (MS).Planta Daninha, v. 10, n. 1–2, p. 25–32.

CLAYTON W.D., RENVOIZE S.A. (1986). Genera graminum. Grasses of the World. [s.l: s.n.].v.13.

CORREIA N.M., DURIGAN J.C. (2009). Manejo químico de plantas adultas de Digitaria insularis com glyphosate isolado e em mistura com chlorimuronethyl ou quizalofop-p tefuril em área de plantio direto. Bragantia, v. 68, n. 3, p. 689–697.

CUNNINGTON J.H., VÁNKY K., SHIVAS R.G. (2005). Lundquistia is a synonym of Sporisorium (*Ustilaginomycetes*). Mycologia Balcanica, v. 2, p. 95–100.

DALIO R.J.D., MAGALHÃES D.M., ATÍLIO L.B., RODRIGUES C.M., BRETON M.C., PICHI S., *et al*. (2014). Efetores na interação planta-patógeno. Revisão Anual de Patologia de Plantas, Passo Fundo, v. 22, p. 25-68.

DODDS P.N., RATHJEN J.P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. Nature reviews. Genetics, v. 11, n. 8, p. 539–48.

DUKE S.O., POWLES S.B. (2008). Glyphosate: a once-in-a-century herbicide. Pest Management Science, v. 64, n. 4, p. 319–325.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (2017). Impacto econômico da resistência de plantas daninhas a herbicidas no Brasil – Circular Técnica 132.

FIGUEIREDO M.B., CARVALHO JUNIOR A.A., GHINI R., BETTIOL W. (1989). Identificação de carvões em capim amargoso *Digitaria insularis* (L.) Mez. em capim colonião (*Panicum maximum* Jacq.), coletados no Estado de São Paulo. In: Arquivos do Instituto Biológico. São Paulo: [s.n.]. p. p.36.

FRANZ J.E. (1985). Discovery, development, and chemistry of glyphosate. The Herbicide Glyphosate.

GEMELLI A., OLIVEIRA JUNIOR R.S. de, CONSTANTIN J., BRAZ G.B.P., JUMES T.M. de C., GHENO E.A., *et al*. (2013). Estratégias para o controle de capim- amargoso (*Digitaria I nsularis*) resistente ao glyphosate na cultura milho safrinha. Revista Brasileira de Herbicidas, v. 12, n. 2, p. 162.

GIRALDO M.C., VALENT B. (2013). Filamentous plant pathogen effectors in action. Nature reviews. Microbiology, v. 11, n. 11, p. 800–14.

GLAZEBROOK J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annual Review of Phytopathology, 43, pp.205-227.

GRUYS K.J., SIKORSLCI J.A. (1998). Inhibitors or Tryptophan, Phenylalanine, and Tyrosine Biosynthesis as Herbicides. p. 371–398.

HOGENHOUT S.A., VAN DER HOORN R.A.L., TERAUCHI R., KAMOUN S. (2009). Emerging Concepts in Effector Biology of Plant-Associated Organisms. MPMI v. 22, n. 2, p. 115–122.

INVASIVE SPECIES COMPENDIUM. https://www.cabi.org/isc/datasheet/109596, 2018.

JAWORSKI E.G. (1972). Mode of action of N-phosphonomethylglycine: inhibition of aromatic amino acid biosynthesis. J Agric Food Chem, v. 20, p. 1195–1198.

JONES J.D.G., DANGL J.L. (2006).The plant immune system. Nature. v. 444 , p. 323.

JONES P. (1999). Control of Loose Smut (*Ustilago nuda* and *U. tritici*) Infections in Barley and Wheat by Foliar Applications of Systemic Fungicides. European Journal of Plant Pathology 105: 729.

KAMOUN S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. Annual Review of Phytopathology 44:41-60.

KISSMANN K.G. (1997). Plantas infestantes e nocivas. São Paulo: Basf Brasileira.

KUVA M.A., FERRAUDO A.S., PITELLI R.A., ALVES P.L.C.A., SALGADO T.P. (2008). Padrões de infestação de comunidades de plantas daninhas no agroecossistema de cana-crua. Planta Daninha, v. 26, n. 3, p. 549–557.

LORENZI H. (1982). Plantas daninhas do Brasil: terrestres, aquáticas, parasitas, tóxicas e medicinais. Nova Odessa: Editora Plantarum.

MACHADO A.F.L., FERREIRA L.R., FERREIRA F.A., FIALHO C.M.T., TUFFI SANTOS L.D., MACHADO M. (2006). Análise de Crescimento de *Digitaria insularis*. Planta Daninha, p. 641–647.

MACHADO A.F.L, MEIRA R.M.S, FERREIRA L.R., FERREIRA F.A., FIALHO C.M.T., TUFFI SANTOS L.D., *et al*. (2008). Caracterização Anatômica de Folha, Colmo e Rizoma de *Digitaria insularis*. Planta Daninha, v. 26, n. 1, p. 1–8.

MEDZHITOV R., JANEWAY C.A. (1997). Innate Immunity: The Virtues of a Nonclonal System of Recognition. Cell, v.. 91, p. 295–298.

MONTEIRO-VITORELLO, C. B., SCHAKER, P. D. C., BENEVENUTO, J., TEIREIRA-SILVA, N., ALMEIDA, S. (2018). "Progress in understanding fungal diseases affecting sugarcane: smut," in Achieving sustainable cultivation of sugarcane, ed. P. Rott (Burleigh Dodds Science Publishing).

OLIVEIRA A.R., FREITAS S.P. (2008). Levantamento fitossociológico de plantas daninhas em áreas de produção de cana-de-açúcar. Planta Daninha, v. 26, n. 1, p. 33–46.

PÉREZ J., Hernández M., Piepenbring M., Minter D.W. (2002). *Sporisorium panici-leucophaei*. [Descriptions of Fungi and Bacteria]. CAB International.

PETERS L.P., CARVALHO G., VILHENA M.B., CRESTE S., AZEVEDO R.A., MONTEIRO-VITORELLO C.B. (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. Planta, v. 245, n. 4, p. 749–764.

PETRE B., SAUNDERS D.G.O., SKLENAR J., LORRAIN C., WIN J., DUPLESSIS S., *et al*. (2015). Candidate effector proteins of the rust pathogen *Melampsora larici- populina* target diverse plant cell compartments. Molecular Plant-Microbe Interactions, v. 28, n. 6, p. 689-700.

PIĄTEK M. (2006). *Sporisorium kenyanum*, a new smut fungus with long twisted sori on *Setaria pallide-fusca* in Kenya. Polish Botanical Journal, v. 51, n. 2, p. 159– 164.

PLISSONNEAU C., BENEVENUTO J., MOHD-ASSAAD N., FOUCHÉ S., HARTMANN F.E., CROLL D. (2017). Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. Frontiers in plant science, v. 8, p. 119.

SHARPEE W.C., DEAN R.A. (2016). Form and function of fungal and oomycete effectors. Fungal Biology Reviews, v. 30, n. 2, p. 62-73.

SELIN C., KIEVIT T.R., BEL MONTE M.F., FERNANDO D.W.G. (2016). Elucidating the role of effectors in plant-fungal interactions: progress and challenges. Frontiers In Microbiology 7: 600.

STOLL M., PIEPENBRING M., BEGEROW D., OBERWINKLER F. (2003). Molecular phylogeny of *Ustilago* and *Sporisorium* species (Basidiomycota, Ustilaginales) based on internal transcribed spacer (ITS) sequences 1.

STOLL M., BEGEROW D., OBERWINKLER F. (2005). Molecular phylogeny of *Ustilago*, *Sporisorium*, and related taxa based on combined analyses of rDNA sequences. Mycological Research, v. 109, n. 3, p. 342–356.

SUNDAR A.R., BARNABAS E.L., MALATHI P., VISWANATHAN R. (2012). A mini-review on smut disease of sugarcane caused by *Sporisorium scitamineum*. InTech Botany, p. 107–128.

TAKANO H.K., de OLIVEIRA R.S., CONSTANTIN J., MANGOLIM C.A., MACHADO M. de F.P.S., BEVILAQUA M.R.R. (2018). Spread of glyphosate-resistant sourgrass (*Digitaria insularis*): Independent selections or merely propagule dissemination? Weed Biology and Management, v. 18, n. 1, p. 50–59.

TANIGUTI L.M., SCHAKER P.D.C., BENEVENUTO J., PETERS L.P., CARVALHO G., PALHARES A., *et al*. (2015). Complete Genome Sequence of *Sporisorium scitamineum* and Biotrophic Interaction Transcriptome with Sugarcane. PLOS ONE, v. 10, n. 6, p. e0129318.

TIMOSSI P.C., DURIGON J.C., LEITE G.J. (2006). Eficácia de glyphosate em plantas de cobertura. Planta Daninha, v. 24, n. 3, p. 475–480.

TSUDA K., KATAGIRI F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. Current Opinion in Plant Biology, v. 13,p. 459-465.

UHSE S. & DJAMEI A. (2018). Effectors of plant-colonizing fungi and beyond. PLoS pathogens, 14(6), e1006992.

UNTERGASSER A., CUTCUTACHE I., KORESSAAR T., YE J., FAIRCLOTH B.C., REMM M., *et al*. (2012). Primer3--new capabilities and interfaces. Nucleic acids research, v. 40, n. 15, p. e115.

VOLLMEISTER E., SCHIPPER K., BAUMANN S., HAAG C., POHLMANN T., STOCK J., *et al*. (2012). Fungal development of the plant pathogen *Ustilago maydis*. FEMS Microbiology Reviews, v. 36, n. 1, p. 59–77.

WEIBERG A., WANG M., FENG-MAO L., HONGWEI Z., ZHIHONG Z., ISGOUHI K., et al. (2013). Fungal Small RNAs Suppress Plant Immunity by Hijacking Host RNA Interference Pathways. Science v. 342, p. 118.

WIN J., CHAPARRO-GARCIA A., BELHAJ K., SAUNDERS D.G.O., YOSHIDA K., DONG S., *et al*. (2012). Effector biology of plant-associated organisms: Concepts and perspectives. Cold Spring Harb Symp Quant Biol. 2012;77:235– 47.

ZABLOTOWICZ R.M., REDDY K.N. (2004). Impact of Glyphosate on the Symbiosis with Glyphosate-Resistant Transgenic Soybean. Journal of Environment Quality, v. 33, n. 3, p. 825.

ZIPFEL C. & FELIX G. (2005). Plants and animals: a different taste for microbes? Current Opinion in Plant Biology, 8(4), pp.353-360.

CHAPTER 2: Smut pathogens: a comparative genomic approach, focusing on the causal agent of sougrass smut disease (*Sporisorium panici-leucophaei*)

ABSTRACT

Fungi from the Ustilaginaceae family are the main causal agents of smut diseases in Poaceae plant species, which compromises crop quality and productivity. Smut species present a biotrophic lifestyle, dependent on an intimate interaction with the host to complete its life cycle and produce spores. The fungus *Sporisorium panici-leucophaei* infects and colonizes invasive sourgrass (*Digitaria insularis*) responsible for the commonly known sourgrass smut disease. Symptomatically, the disease resembles another one caused by *S. scitamineum* in sugarcane, producing a whip-like structure from the apex of the plants where sporogenesis take place. Despite similarities (life cycle and symptom induction) with sugarcane smut little is known about this pathosystem and whip induction. Therefore, we proposed to sequence the entire genome of *S. panici-leucophaei* and compare to that of *S. scitamineum* searching for species-specific genes and those shared between them that may help to explain similarities in the symptom development. In addition, experimental data concerning the pathogen life cycle and specific fungal structures required for plant infection were objectified. The characterization of *S. panici-leucophaei* life cycle *in vitro* was concluded with the distinction of fungal phases, such as spore germination, probasidium formation, sporidia formation and dikaryotic hyphae visualization. Moreover, we quantitatively established the *in planta i*nfection *time-course* by spore germination 8 hours post-inoculation (hpi); appressorium formation at 48 hpi and hyphae detection at 72 hp. The genome of the haploid sporidia (Spl10) was sequenced using the Illumina and Oxford Nanopore platforms. The complete assembly equence of *S. panici-leucophaei* comprised of 18,913,471 nucleotides distributed in 22 contigs (plus a mtDNA sequence). We used Augustus to predict 6,404 genes and compare them against other smut genomes, available in public databases. OrthoMCL results showed a total of 7,277 ortho-groups described among smut species, whereas, S. *panici-leucophaei*, from a total 6,404 genes, 5,955 were considered single-copy and 449 paralogs. Pathogens were screened for candidate effector genes, where both present a similar effector repertoire (17% of all secreted proteins). Only members of the *Mig2* Family were not found in *S. panici-leucophaei*, in comparison with the sugarcane smut. The results yielded the first insights into shared and species gene-content and provided useful information on a poorly studied plant pathogen, opening novel research opportunities concerning this pathosystem.

Keywords: Sourgrass, Smut fungi, Effector biology, Next Generation Sequencing (NGS)

1.) Introduction

Fungi of the family Ustilaginaceae are the primary causal agents of smut diseases in plants of the Poaceae family. The best-known genera are *Ustilago* and *Sporisorium*, which infect species of agronomic importance, such as cereals, forages, and sugarcane (Horst, 2013). Smut

diseases occur in most of the producing countries of their respective hosts (Begerow *et al*., 2014). Losses range from negligible to severe proportions, depending on the variety and environmental conditions (Sundar *et al*., 2012; Diaz-Quijano *et al*., 2016).

The smut species are biotrophic fungi, dependent on an intimate interaction with the host to complete its life cycle. The main symptom of smut diseases is the formation of a black mass of spores resembling a soot (Sundar *et al*., 2012). Under favorable environmental conditions, the spores germinate on the plant surface and form haploid sporidia. The haploid sporidia present a mating system in which a compatible sexual reaction requires different mating-types (Bakkeren & Schirawski, 2008). Compatible mates can fuse, forming the infective dikaryotic hyphae capable of penetrating the host tissue. After penetrating plant tissues, colonization becomes systemic, and intense proliferation occurs close to the meristematic region of the plant (Carvalho *et al*., 2016).

Although smut species are phylogenetically close (Stoll *et al*., 2003; Stoll *et al*., 2005), these species differ in the range of hosts and site of infection, *in planta* proliferation, and symptom development and progression. Most smut species infect only a few hosts or are restricted to a single plant species (Begerow *et al*., 2004). Host specialization processes involve host-specific adaptations, which make the pathogen capable of infecting, overcoming defense mechanisms, capturing nutrients, multiplying, and reproducing within the new host (Haueisen & Stukenbrock, 2016). However, there is no consensus on what determines the pathogen-host range. The process of host specialization is still poorly understood and probably involves different evolutionary histories and different molecular mechanisms in each phytopathogen (Kirzinger & Stavrinides, 2012).

Recently, Benevenuto *et al*. (2018) revealed significant similarity in the protein pool, encoded by the genomes of nine fungal species that cause smut on distinct hosts. Seven were available in databases, and two new genomes were sequenced, assembled, and annotated in the study. The species were isolated from eight distinct hosts, five of which are of agronomic importance and three from non-domesticated plants: *Ustilago maydis* (corn), *Sporisorium reilianum* (corn); two isolates of *Ustilago hordei* (barley and oats); *Ustilago tritici* (wheat); *Sporisorium scitamineum* (sugarcane); *Ustilago esculenta* (*Zizania latifoli*a, a wild relative of rice); *Ustilago trichophor*a (*Echinochloa colona*, a species of grass); *Melanopsichium pennsylvanicum* (*Persicaria* sp., a dicotyledonous species). Among the genes identified, candidate effector genes stand out as potential determinants of host specialization.

Effectors genes are secreted molecules that act by modulating host cell physiological and morphological processes (Dalio *et al*., 2014; Selin *et al*., 2016), enabling pathogen colonization

within plant tissues. Therefore, *in silico* identification and characterization of these candidate effector genes is a crucial step to understand common and specific pathogenicity mechanisms among microorganisms.

Smut core effectors have been functionally characterized over the past few years, mainly on the maize smuts *U. maydis* and *S. reilianum* (Lanver *et al*., 2017). Effector genes encoding proteins like Pit2 (*Protein Involved in Tumors 2*) and Stp1 (*Stop After Penetration 1*) act as cysteine protease inhibitors (Doehlemann *et al*., 2011; Liang, 2013); whereas See1 (*Seedling Efficient Effector 1*) and Cmu1 (*Chorismate Mutase-1*) influence cell division related tumorigenesis and chorismate mutase pathways, respectively (Djamei *et al*., 2011; Doehlemann *et al*., 2011). Pep1 (*Protein essential during penetration-1*) delays the production of ROS and the activation of the plant defense response (Hemetsberger *et al*., 2012). Tin2 (*Tumor Inducing 2*) is a cytoplasmic effector that acts influencing anthocyanin biosynthesis, while inside the host (Tanaka *et al*., 2014, Brefort *et al*., 2014).

Extending comparative genomics analyses of smut species should help to understand the evolution of these pathogens and contribute to defining mechanisms of pathogenicity, especially by identifying specific candidate effector genes.

Recently, a sourgrass smut pathogen was detected and isolated from a whip-like structure from a sourgrass plant in a sugarcane field, near Ilha Solteira in São Paulo State (Brazil). The incidence of smut in sourgrass (*Digitaria insularis*) caused by *Sporisorium panici-leucophaei* (= *Lundquistia panici-leucophaei*) was described for the first time in 1989 between the lines of a coffee plantation in Itu, also in São Paulo State (Brazil) (Guini *et al*., 1990). *S. panici-leucophaei* shows a similar life cycle as described to other smut fungi.

The sourgrass (*Digitaria insularis* (L.) Fedde) is a plant species native to tropical and subtropical regions often found in pastures, coffee plantations, orchards, and agricultural areas, from the Asian to the American continent (Mondo *et al*., 2006; Machado *et al*., 2008). Sourgrass presents high production of seeds coupled with enormous germinative power, features that provide raised wind dissemination capacity. Moreover, *D. insularis* has a fast and aggressive development life cycle, reproducing both by seeds and rhizomes (Kissmann, 1997).

In addition to *D. insularis* vast geographical dispersal and high competitiveness, sourgrass has been gaining importance in recent years due to its high tolerance to chemical control based on the use of Glyphosate, a notable herbicide used worldwide. Since the development of the whip-like structure is a hallmark of the sugarcane smut disease, and a similar structure is formed during *S. panici-leucophaei* and sourgrass interaction, we hypothesized that similar

mechanisms of pathogenicity and symptom development might exist between these species. Through complete genome sequencing of the *S. panici-leucophaei* and comparative genomic analyses, we can achieve first insights on *S. panici-leucophaei* shared and species-specific gene content and also identify its core set of candidate effector genes by *in silico* prediction.

2.) Materials and Methods

2.1.) Biological Material

Smut-infected *Digitaria insularis* plants were collected in the fields of the Fazenda de Ensino, Pesquisa e Extensão – FEPE / UNESP – Câmpus de Ilha Solteira, located in the city of Selvíria, Mato Grosso do Sul State, Brazil (Latitude: 20°25′58″ S; Longitude: 51°20′33″ O) Samples were kindly provided by Prof. Paulo Ceresini (State University of Ilha Solteira) (Fig. 1). No special permits were necessary for whip and teliospores collection and use. This work does not involve endangered or protected species.

Figure 1. Smut-infected *D. insularis* collected in the Fazenda de Ensino, Pesquisa e Extensão – FEPE / UNESP – Câmpus de Ilha Solteira, located in the city of Selvíria, Mato Grosso do Sul State, Brazil**.** Red arrows indicate the whip-like structures where sporogenesis occurs. Ruler used as a comparison measure (= 20 cm).

Diploid teliospores were collected from whips and transferred to microtubes containing distilled water added of the antibiotics: streptomycin sulfate (5 mg/L) and tetracycline (2.5 mg/L) and incubated for 30 minutes at room temperature. Three serial dilutions were

performed, 100 µL of each suspension was plated in solid YM medium (3 g/L of yeast extract; 3 g/L of malt extract; 5 g/L of soybean peptone; 10 g of glucose; 15 g/L of agar for solid medium), and incubated overnight at 28 °C in the dark. After 12 hours, teliospores germination and hyphal growth were observed under a light microscope. Hyphal fragments of approximately 2 cm were transferred to YM liquid medium and incubated in an orbital shaker at 28 °C to obtain the yeast-like haploid cells (haploid sporidia resulting from clonal growth after meiosis). Yeast-like haploid cells were diluted, plated on YM medium, and isolated for mating compatibility tests. The pairwise mating compatibility was analyzed in a mating-plate experiment (Alexander & Srinivasan, 1966; Bölker *et al*., 1992). Compatible mating cells were identified by hyphal development upon mating reaction, according to Carvalho *et al*. (2016). We selected two haploid compatible strains that originated from the SPL10 initial growth for further experiments.

The seeds of *D. insularis* plants used in the experiments were kindly provided by Prof. Pedro Jacob Christoffoleti (University of São Paulo). Seeds were stored at -4 ºC for further use.

2.2.) Interspecific hybridization assay for mating-type identification

Haploid strains with opposite mating-types derived from the SPL10 colony were paired with either SSC04 haploid strains A or B of *Sporisorium scitamineum* according to the protocol of Carvalho *et al*. (2016). Compatible strains with either SSC04 reference strains A and B as determined by the development of white aerial hyphae assigned mating-types A or B for the fungal strains. The compatible haploid strains of each isolate (SPL10 and SSC04) were separately inoculated into liquid YM medium and incubated at 28 °C overnight under constant agitation (200 rpm). For the interspecific hybridization test, 5 µL of each suspension (Optical Density of 1 at 600 nm) were inoculated at the same reaction site in a Petri dish containing solid YM culture medium (Alexander & Ramakrishnan, 1980; Bauch, 1923). All haploid yeast-like cells were crossed against each other in order to cover all matting possibilities, as described in Singh *et al*. (2004). Hyphae fusion from compatible yeast-like cells indicates the presence of sexual reaction and thus the formation of visually recognizable interspecific hyphae (formation of fluffy white hyphae at the reaction site). The tests were performed with five biological replicates for each combination (five Petri dishes) containing 10 reaction sites on each plate.

2.3.) Characterization of *S. panici-leucophaei* morphological structures

The main stages part of the smut fungi life cycles (Teliospore, Germinated Teliospore, Yeast-like Phase, and Dikaryotic Phase) were identified for *S. panici-leucophaei* based on Sundar *et al*. (2012). Each of the typical structures was individually analyzed using LEO 435 VP Scanning Electron Microscopy at Support Center for Electron Microscopy Research Applied to Agricultural Research (NAP/MEPA) from ESALQ/USP, whose preparation methods are detailed below.

2.3.1.) Teliospores and Germinated Teliospores

Teliospores were spread over the surface of a water-agar medium (7.5%) and incubated at 28 ºC for 24 hours to stimulate spore germination. For teliospores observations, they were spread over plates of the water-agar medium not incubated and immediately prepared and fixed according to Roland (1978). Both samples were processed and deposited on a coverslip for fixation with osmium tetroxide vapor. This coverslip was placed in a petri dish with damp filter paper to form a wet chamber. In the fume hood, two drops of 1% osmium tetroxide were placed in the microtube lid and placed in the Petri dish. They were then wrapped in aluminum foil and left in the fume hood for 24 hours. Then, the material was metalized using Baltec SCD 050 Sputter and visualized in Scanning Electron Microscopy. Spore measurements were performed by comparing 50 spores with the Scanning Electron Microscopy scale bar.

2.3.2.) Yeast-like Phase

In solid YM medium, yeast-like cells were spread and incubated for four days at 28 °C. Then, a single colony was transferred to liquid YM medium and incubated with shaking at 28 °C until reaching an OD (Optical Density) of approximately 1. The preparation included applying 0.1% poly-L-lysine following incubation for 15 min at room temperature. The removal of the excess of poly-L-lysine immediately followed the application of a drop of liquid medium containing *S. panici-leucophaei* cells. The slide was placed in a Becker with modified Karnovsky fixative (2.5% Glutaraldehyde, 2.5% formaldehyde in 0.05 M sodium cacodylate buffer pH 7.2, 0.001 M CaCl2) for 24 hours. Then, dehydration was applied with ethanol 30, 50, 70, 90 and three times 100% (Horridge & Tamm, 1969). At the end of the dehydration, the slide was taken to Baltec SCD 030 critical drying device.

2.3.3.) Hyphal Phase

The fungus was inoculated in solid YM culture medium and incubated for one week at 28 °C. Hyphae preparation followed as applied for the Teliospore and Germinated Teliospore phases, mentioned above.

2.4.) *In planta* infection time course

D. insularis seeds were grown in pots of 5 L containing the substrate "Basaplant - Vegetables" for 10 days in greenhouse conditions. Leaves of 10 days-old plants were detached, incubated *overnight* over controlled humidity and drop inoculated with 20 uL of a suspension containing 5 x 10^6 teliospores mL-1 in 0.01% (v/v) Tween-20, whereas controls were inoculated with 20 uL of sterile saline solution (0.85%) in 0.01% (v/v) Tween-20 (mock). The inoculated leaves were maintained in a wet chamber and incubated at 28° C. Leaves were examined 6, 8, 24, 48, and 72 hours post-inoculation (hpi) for the stage of fungal growth. We detected fungal structures by using a lactophenol-cotton blue stain (10 g phenol, 10 mL glycerol, 10 mL lactic acid, 0.02 g cotton blue, and 10 mL deionized water) (Tuite, 1969). We used the Light microscope Optika B-350 (Optikam B5 digital camera) to analyze all-time points post-inoculation. The experiment was performed with three biological replicates. We determined the percentage of germinated spores at 6 and 8 hpi by counting the number of it in a total of 100 spores for each replicate (James, 1973). We then inspected at least 100 germinated spores for appressorium formation for each replicate at 24 and 48 hpi (Apoga *et al*., 2004, Peters *et al*., 2017). The significance of the observed differences was verified using Student's test ($P < 0.05$).

2.5.) *S. panici-leucophaei* quantification (qPCR-assay) in *D. insularis*

D. insularis seeds were grown in pots of 5 L containing the substrate "Basaplant - Vegetables" in a randomized experimental design for 10 days (after seed germination) in greenhouse conditions. After that, plants were individually transplanted to pots of 5 L containing the same substrate mentioned before and maintained under greenhouse conditions for 2 days. Leaves were drop inoculated and punctured with 20 uL of a suspension containing 5 x 10^9 teliospores mL-1 of *S. panici-leucophaei*. Entire plants were collected at 8 hpi, 48 hpi and 10 days post-inoculation (dpi) and taken to liquid nitrogen immediately for further analysis. Each time-point consisted of plant material from 5 biological replicates (each replicate consisted of a *pool* of 3 entire plants).

Plant material was macerated using liquid nitrogen, weighed and stored at -80 ºC. 50 milligrams of plant material from each biological replicate were used for DNA extraction, performed with the modified CTAB method (Porebski *et al*., 1997). Primers designed to target the Internal Transcribed Spacer (ITS) region of *S. panici-leucophaei* genome (F 5' TTCGCTCTTTCTTCCCTGCC 3' and R 5' AGCTCGTACCCTTTCTTCGC 3') were used for

the qPCR assay. DNA quantification (150 bp amplicon) was performed on a Fast Real-Time 7300 PCR System (Applied Biosystems) using the GoTaq® qPCR Kit and RT-qPCR Systems - Promega ®. The conditions of qPCR cycles comprised 2 minutes at 95 °C, 95 °C for 15 seconds until denaturation and 1 minute at 60 ºC for annealing and extension (40 times).

A standard curve was generated by 10-fold dilution series (1000 ng to 1 pg) of SplA DNA. The initial input of target DNA (infected plant) in each reaction was 100 ng/µL. We estimated the concentration of fungal DNA in infected plants at 8 hpi, 48 hpi and 10 dpi.

2.6.) Fungal DNA extraction

Genomic DNA of haploid cells derived from a unique mating-type (SPL10A) growing in liquid YM medium was used for PCR reactions, fungal DNA quantification, and complete genome sequencing. DNA was extracted using the "Genomic-tip 20G" (Qiagen, Inc.) kit, according to the manufacturer's instructions for yeast-like cells. DNA integrity was checked on agarose gel and quantity measured using a Qubit 2.0 Fluorometer (Invitrogen, ThermoFisher Scientific, Waltham, MA USA).

2.7.) ITS-based phylogeny

The phylogeny of ten different smut species was constructed based on Internal Transcribed Spacer (ITS) DNA sequences (see Suppl. Table 1) recovered from GenBank using the maximum likelihood method in the MEGA V 6.0 software (Tamura *et al*., 2013). The ITS of *S. panici-leucophaei* was isolated through PCR amplification using the universal ITS primers (ITS1 Forward 5' GGT GAA CCT GCA GAT GGA TC 3') and (ITS4 Reverse 5' TCC TCC GCT TAT TGA TAT GC 3') (Stoll *et al*., 2003). The amplicon was purified using llustra™ GFX™ PCR and further the DNA sequencing (Sanger method) performed at the GaTELab laboratory of Prof. Dr. Marie-Anne Van Sluys (IB/USP).

2.8.) Whole-genome sequencing strategies

S. panici-leucophaei genome was sequenced, combining two strategies: Nanopore Oxford and Illumina technologies.

For Illumina sequencing, the whole-cell Spl10A DNA library was prepared using the Nextera DNA Sample Preparation Workflow (Illumina paired-end) kit. Paired-end libraries

were sequenced using Illumina HiSeq 2000. The sequencing was performed at the Helixxa Serviços Genômicos S.A.A total of 715,363 reads of 300 bp in length (approximately 430 Mbp) corresponding to more than tenfold of sampling depth was analyzed using FASTQC (ver. 0.10.1) and Seqyclean (ver. 1.8.10) to filter low-quality sequences (Phred quality below 20).

For the Oxford Nanopore platform, the Genomic DNA by Ligation (SQK-LSK109) kit was used for library preparation. A total amount of 1 ug of high molecular weight DNA in 48 uL was used for NEBNext FFPE DNA Repair (New England Biolabs) in a total volume of 60 uL including 3.5 uL NEBNext FFPE DNA Repair Buffer, 2 uL NEBNext FFPE DNA Repair Mix, 3.5 uL Ultra II End-prep reaction buffer and 3 uL Ultra II End-prep enzyme mix. For the Adapter Ligation and Clean-up step, 60 uL of eluate from the previous step was used in a total volume of 100 uL containing 25 uL Ligation Buffer (LNB), 10 uL NEBNext Quick T4 DNA Ligase and 5 uL Adapter Mix (AMX). The final library was eluted in 15 uL of Oxford Nanopore Elution Buffer (Oxford Nanopore Technologies), and 1 uL was used for Qubit dsDNA BR (ThermoFisher Scientific) quantification.

Following the manufacturer recommendation, 5-50 fmol of the final library elution was used for loading the flow cell. Sequencing was performed on MinION Flow Cells (R9.4.1) (Oxford Nanopore Technologies; catalog no. FLO-MIN106_SQK). Immediately before the start of the sequencing run, a Platform QC was performed to determine the number of active pores. Priming of the flow cell was performed by applying 800 uL priming buffer through the sample port, and after 5 minutes incubation at room temperature, 200 uL of priming buffer was loaded again through the sample port with opened SpotON port. After that, 12 uL of the final library was mixed with 25.5 uL Loading Beads (Oxford Nanopore Technologies LB) and 37.5 uL Sequencing Buffer (Oxford Nanopore Technologies SQB). Directly after priming, 75 uL of the prepared library was loaded through the SpotON port. Base-calling was performed simultaneously upon completion of the sequencing run. A total of 830 Mbp was obtained consisting of 129,968 reads, the largest one obtained of 96,286 bp, and an average of 7,52 Kbp read size. Reads under 1,000 bp were not used in the assembling.

2.9.) Genome Assembly

De novo sequence assembly was performed applying a hybrid strategy using Illumina and Oxford Nanopore reads. For that, SPADES (v3.13.1) was used with default parameters for Illumina reads and Canu 1.8 (Koren *et al*., 2017) was implemented using default parameters for nanopore reads (-pacbio-raw) and estimated genome size of 20 Mbp. Cross_match (www.phrap.org) was used to close gaps in the scaffolds, generating an assembly (v3) with 22

Chromosomes. Telomeric repeats were manually analyzed, searching for the canonical repeated sequence TTAGG (Benevenuto *et al.*, 2016). First, Illumina reads that aligned with *S. scitamineum* genome were assembled generating a draft version (v1). A second version (v2) was assembled with Nanopore reads that aligned to the first version (v1). The complete genome sequence of *S. panici-leucophaei* (v3) was generated by implementing Pilon (v1.18) to v2 in order to correct indels using Illumina reads.

BWA-MEM was used to align Illumina short reads to nanopore assembly in order to use Pilon v1.18 (Walker *et al*., 2014) to polish the assembly, i.e., fix indels that are sequencing artifacts of nanopore technique and close possible present gaps (parameter --fix bases, gaps). With RepeatModeler v. 1.0.11 and repeatmasker v. 4.1.0, we identified repetitive regions in the genome. BUSCO *v3* (Waterhouse *et al*., 2017) was used to evaluate the completeness of the genome assembly, using a Basidiomycota database as reference.

2.10.) Genome Annotation

Gene predictions were obtained by using Augustus (v3.3.3, parameter --species=ustilago_maydis) with proteins of *S. scitamineum* (Taniguti *et al*., 2015) as extrinsic evidence, as described in section "alternate protocol 7" of Hoff & Stanke (2019).

2.11.) Predicted Secretome and Effectorome

The programs SignalP version 5.0b (Armenteros *et al*., 2019), TMHMM V2.0c (Krogh *et al*., 2001) and PS-Scan (de Castro *et al*., 2006) were used to predict those proteins that present a signal peptide, does not present a transmembrane domain and does not present a ER retention signal, respectively. This set of proteins was considered potentially secreted and defined as the fungus predicted secretome. Also, LOCALIZER 1.0.4 (Sperschneider *et al*., 2017) was used to predict the subcellular localization of this set of proteins in plant cells. EffectorP 2.0 was used to predict the candidate effector repertoire (Effectorome) from the predicted secretome based on machine learning (Sperschneider *et al*., 2016). Characterized effectors in smut species (core effectors) were screened for orthologs using OrthoMCL and tblastN.

The predicted proteome was also screened for CAZymes (carbohydrate-active enzymes) (Lombard *et al*., 2013) using Hmmscan from the HMMER v3.1b2 package4 and the dbCAN HMM profile database (Yin *et al*., 2012).

2.12.) Orthologs group analysis

For orthology and paralogy determination, ten smut (Ustilaginomycotina) genomes available in public databases were used: *S. panici-leucophaei* (Spl)*, Ustilago hordei* (UhoO)*, Ustilago hordei* (UhoB) *Ustilago tritici syn. Tilletia tritici* (Utri)*, Ustilago maydis* (Umay)*, Ustilago esculenta* (Uesc)*, Ustilago trichophora* (Utcp)*, Sporisorium reilianum* (Srei)*, Sporisorium scitamineum* (Ssci) *and Melanopsichium pennsylvanicum* (Mpen) (Table 1).

Orthologous and paralogous groups among the ten genomes were determined using OrthoMCL with default parameters: BLASTp e -value cutoff of 1e-5, percent match cutoff of 50, and inflation index of 1.5 (Li et *al*., 2003). The output of OrthoMCL used to discriminate among core and unique clusters, single-copy, singletons, co-orthologs groups and paralogs genes.

TABLE 1 | List of analyzed Ustilaginomycotina species, strains and genomes assemblies.

	Abbreviations	Species	Strain	Host/source	Project number	Reference
SMUTS	Spl	*Sporisorium panici-leucophaei*	SPL10A	*Digitaria insularis (sourgrass)*	Submitted	This work
	UhoO	*Ustilago hordei*	Uhor01	*Avena sativa (oats)*	PRJNA393983	Benevenuto et al. 2018
	UhoB	*Ustilago hordei*	Uh4857-4	*Hordeum vulgare (barley)*	PRJEA79049	Laurie et al. 2012
	Utri	*Ustilago tritici syn. Tilletia tritici*	CBS119.19	*Triticum spp. (wheat)*	PRJNA400640	Benevenuto et al. 2018
	Umay	*Ustilago maydis*	521	*Zea mays (maize)*	PRJNA1446	Kamper et al. 2006
	Uesc	*Ustilago esculenta*	MMT	*Zizania latifolia (rice-relative)*	PRJNA263330	Ye et al. 2017
	Utcp	*Ustilago trichophora*	RK089	*Echinochloa colona (wild grass)*	PRJNA316802	Zambanini et al. 2016
	Srei	*Sporisorium reilianum*	SRZ2	*Zea mays (maize)*	PRJNA64587	Schirawaski et al. 2010
	Ssci	*Sporisorium scitamineum*	SSC39B	*Saccharum spp. (sugarcane)*	PRJNA275631	Taniguti et al. 2015
	Mpen	*Melanopsichium pennsylvanicum*	Mp4	*Persicaria sp. (wild dicot plant)*	PRJEB4565	Sharma et al. 2014

3.) Results

3.1.) Structures of *S. panici-leucophaei*

The distinct phases of *S. panici-leucophaei* development were identified during the development of the fungus *in vitro*. Each phase presented compatible features with the smut pathogen strategy at each moment of the fungus biological life cycle. Diploid teliospores (Fig. 2A), structures of resistance of the pathogen to abiotic stresses, were identified and showed similarity to the spores described by Vanky (2005). The author demonstrated that such teliospores have a variable length between 6.0 - 9.5 µM, compatible with the measurement performed in this study (6.0 - 7.0 µM). Other features of *S. panici-leucophaei* spores are their variable shapes and echinulation (Piatek, 2006). Under specific humidity and temperature conditions, the spores germinate (Fig. 2B) developing probasidium, a structure that will later form the sporidia, commonly described as haploid yeast-like cells. This elongated structure undergoes meiosis forming four yeast-like cells (Fig. 2C), of which two sexually opposite and compatible sporidia will emit reproductive conjugation hyphae for sexual reaction. The formation of four yeast cells from the probasidium was not identified in this study by Scanning Electron Microscopy.

The dikaryotic hyphae (Fig. 2D), which originated from the sexual reaction between two opposite and compatible sporidia, will produce appressorium, a structure responsible for

penetration of the host tissues. After penetration, the pathogen will continue as dikaryotic hyphae in order to colonize the host's tissues. Finally, to complete its biological cycle, *Sporisorium panici-leucophaei* modulates the physiology and morphology of its host, in order to emit a structure composed of plant tissue and also their own fungal tissues, commonly known as the whip (Fig. 1). This structure consists of the place where sporogenesis occurs and, therefore, the region from which such resistance structures will be dispersed, restarting the disease cycle again.

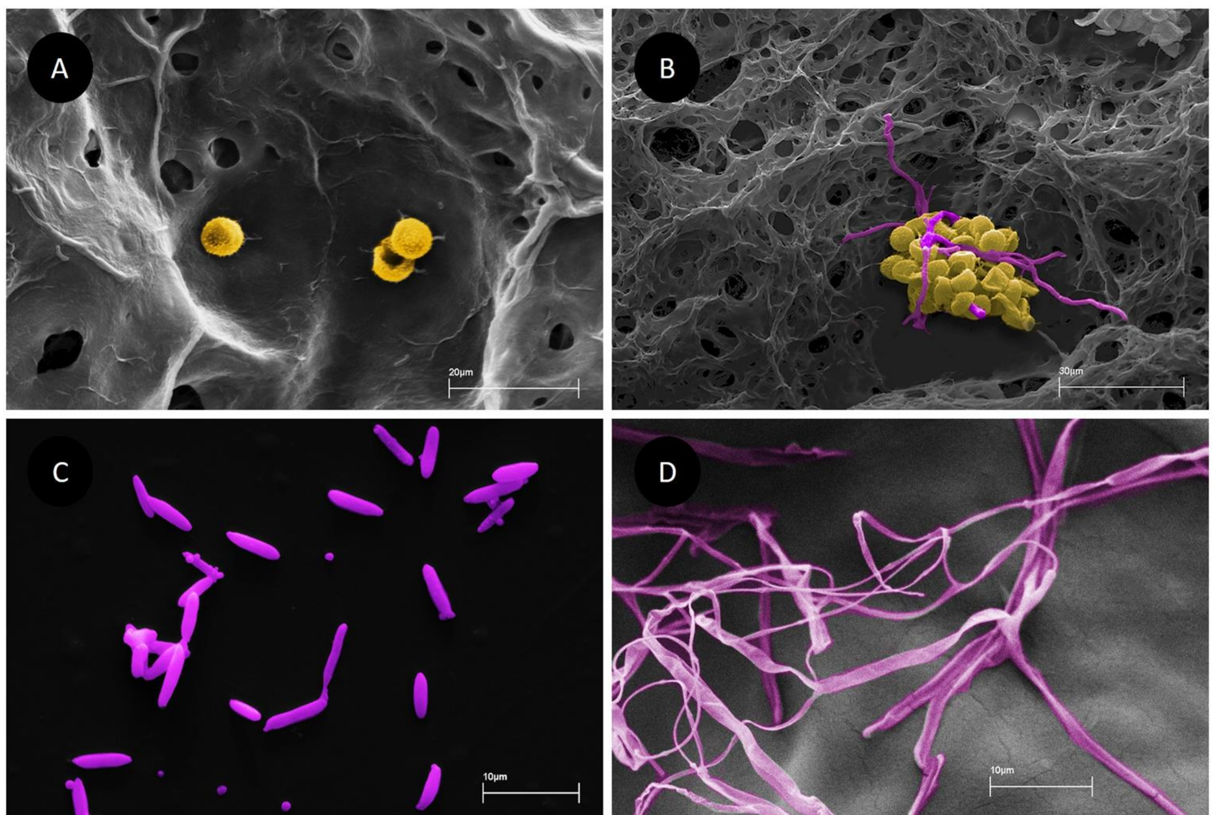

Figure 2. *Sporisorium panici-leucophaei* in culture medium recorded by scanning electron microscopy. (A) Non-germinated teliospore adhered to the agar-water medium (7.5%). (B) Set of germinated teliospores emitting probasidium. (C) Yeast-like haploid cells in liquid YM culture medium. (D) Dikaryotic hyphae growing in solid YM culture medium. Image artificially colored with Adobe® Photoshop software.

The sporogenesis process and, consequently, the emission of the whip is a host-dependent event in which the pathogen necessarily needs to be located *in planta* and not *in vitro*. Structures associated with fungal penetration and tissue colonization were identified in the time-course experiment where the fungal structures were stained with lactophenol-cotton blue. *D. insularis* leaves were used to quantify spore germination at 6 and 8 hpi, while appressorium formation was visualized at 24 and 48 hpi, and tissue colonization observed at 72 hpi.

At 6 hpi, 40% of teliospores germinated; however, at 8 hpi, teliospore germination reached a maximum rate (78%) (Fig. 3A). Appressorium formation was quantified by observing fungal filaments on leave surfaces of at least 100 germinated spores. 12% of the hyphal tips developed an appressorium at 24 hpi (Fig. 3B), whereas at 48 hpi, 52% developed an appressorium. An extensive network of filaments, commonly known as hyphae, was observed at 72 hpi.

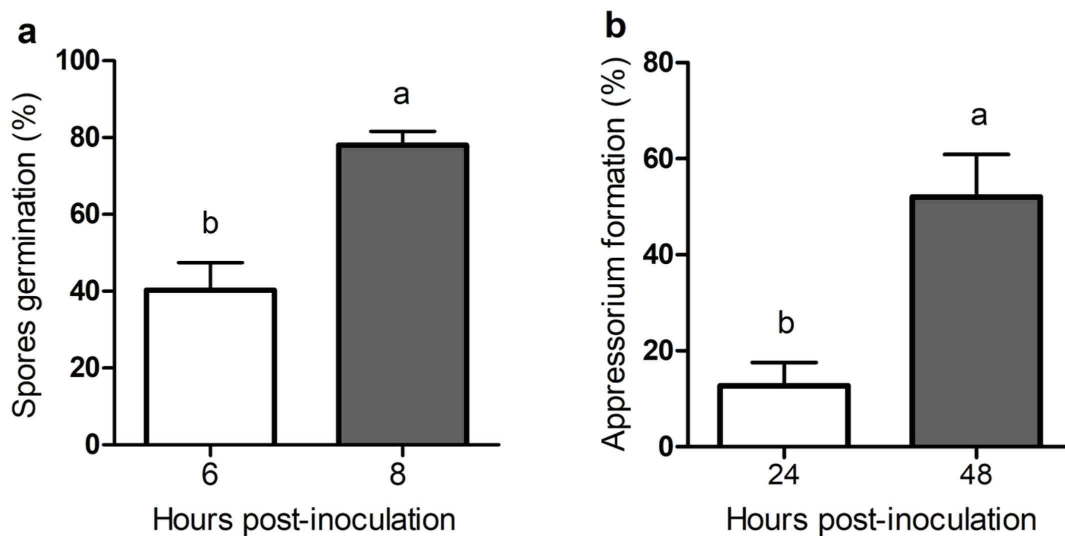

Figure 3. Teliospore germination and appressorium formation of *S. panici-leucophaei* on sourgrass leave surface. (a) Teliospore germination (%) at 6 and 8 hours post-inoculation (hpi). (b) Appressorium formation (%) at 24 and 48 hpi. Bars represent the standard errors of three independent biological replicates. Different letters represent statistically significant differences ($P < 0.05$) between time-points.

3.2.) *S. panici-leucophaei* growth *in planta*

After characterizing *S. panici-leucophaei* structures associated with penetration and tissue colonization, we quantified fungal growth in *D. insularis* plants at three time-points: spore germination (8 hpi), appressorium formation (48 hpi) and hyphae colonization (10 dpi).

This Sybr green-based quantification assay was the first attempt to quantify the sourgrass smut fungi. We used the ITS instead of Intergenic spacer (IGS), as previously described for *S. scitamineum* (Peters *et al*., 2017). The resulting standard curve showed a linear correlation between Ct values (cycle threshold) and *S. panici-leucophaei* DNA amounts, described by the equation y = -3.450828x + 32,92 and R2 = 0.99979. The standard curve was stable, detecting a range from 952.98 ng of fungal DNA to 0.0098 ng (Fig. 4).

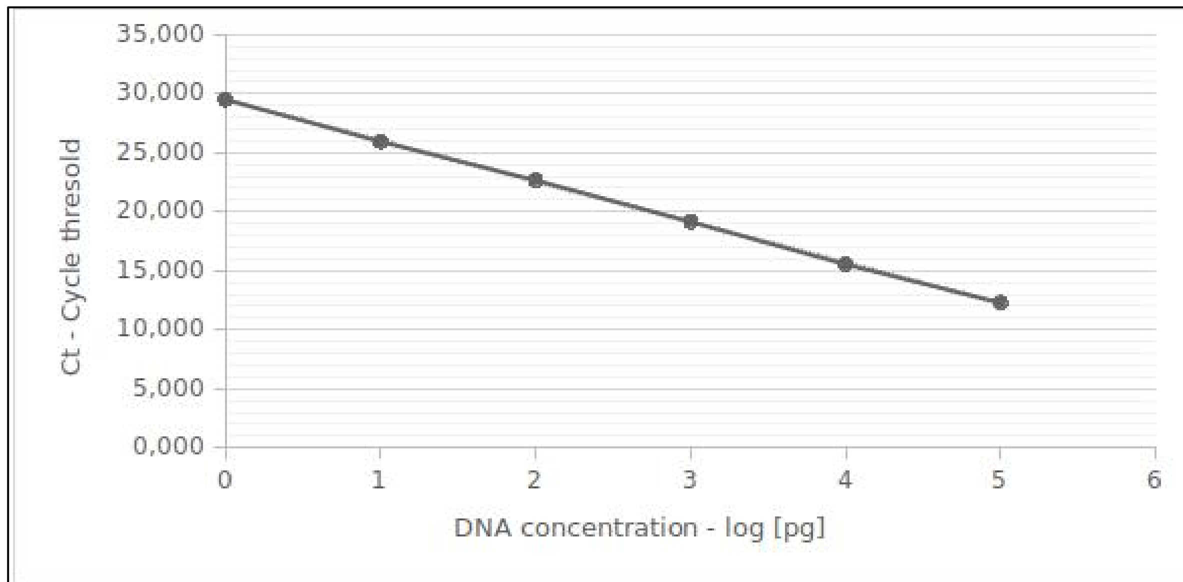

Figure 4. Standard curve of a linear correlation between Ct values and *S. panici-leucophaei* quantification. The equation of the standard curve is y = -3.450828x + 32,92 and R2 = 0.99979.

We could detect fungal DNA inside the plant in all time-points analyzed. However, the amount of fungal DNA in the sample taken, reduced instead of increase, as expected (Fig. 5). This apparent reduction is probably due to the faster growth of the plant compared to the pathogen. Because we did not have apparent symptoms at the time of collection, the whole plant in each time-point was collected. The ideal quantification would consider the amount of plant cell duplication, and initiate the experiment using the same amount of plant before the maceration of the tissue. Also, at earlier time-points, we may be quantifying the DNA of ungerminated teliospores.

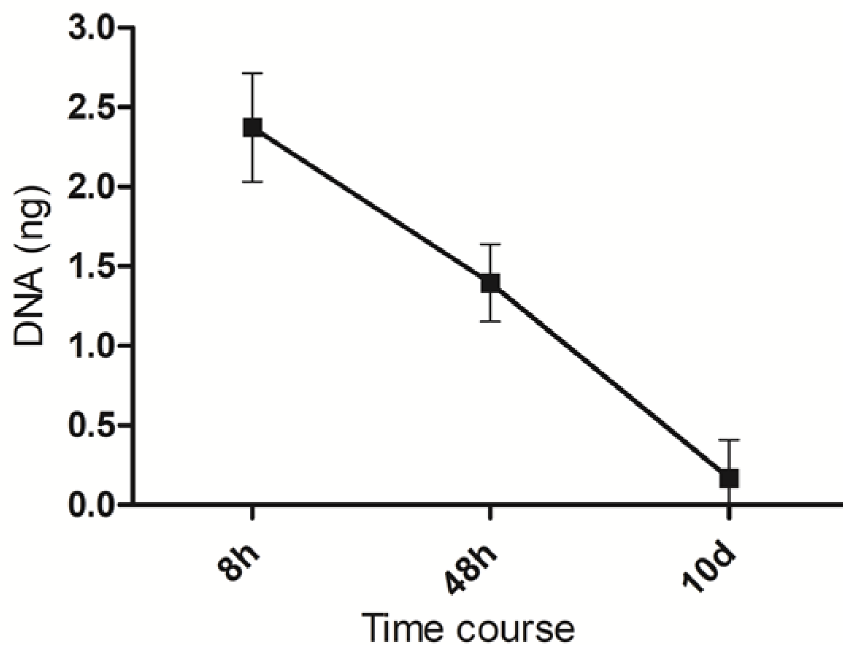

Figure 5. DNA quantification of *S. panici-leucophaei* during *D. insularis* infection process by qPCR. Infection time course at 8h, 48h and 10d. Bars represent the standard error.

3.3.) Interspecific Hybridization Assay

The dimension of interspecies sexual compatibility within smuts is still unclear, despite studies pointing to a high hybridization compatibility among them (Kellner *et al*., 2011). Therefore, to verify whether *S. panici-leucophaei* and *S. scitamineum* are potentially mating partners, we performed an interspecific hybridization test. Firstly, we separated opposite and compatible haploid yeast-like cells from each isolate (Spl10 A/B and SSC04 A/B) and crossed them against each other in order to cover all mating possibilities. After that, compatible interspecific hyphae among the mating tests were screened with the aid of a light microscope.

Fifty single mating tests were performed comprising five replicates of 10 different mating tests (reaction sites) under the conditions mentioned above. Crosses between opposite cell types within each species (Spl0 A *vs.* Spl0 B and Ssci A *vs.* Ssci B) were used as positive controls. Therefore, matings between equal cell types (Spl0 A *vs.* Spl0 A, Spl0 B v*s.* Spl0 B and Ssci A *vs.* Ssci A, Ssci B *vs.* Ssci B) were designated as negative controls.

As expected, all positive controls produced visually recognizable hyphae with formation of fluffy white filaments at the reaction site (Fig. 6A and 6B). As foreseen again, all mating tests characterized as negative controls did not form filaments. Interestingly, although no aerial whitish hyphae structure was visualized, all mating tests between opposite cell types of interspecific confrontations (Spl A *vs* Ssci B and Spl B *vs* Ssci A) showed conjugation hyphae

and consequently filament formation at light microscopy (Fig. 6C). Interspecific confrontations with equal cell types (Spl A *vs* Ssci A and Spl B *vs* Ssci B), as expected, were not sexually compatible.

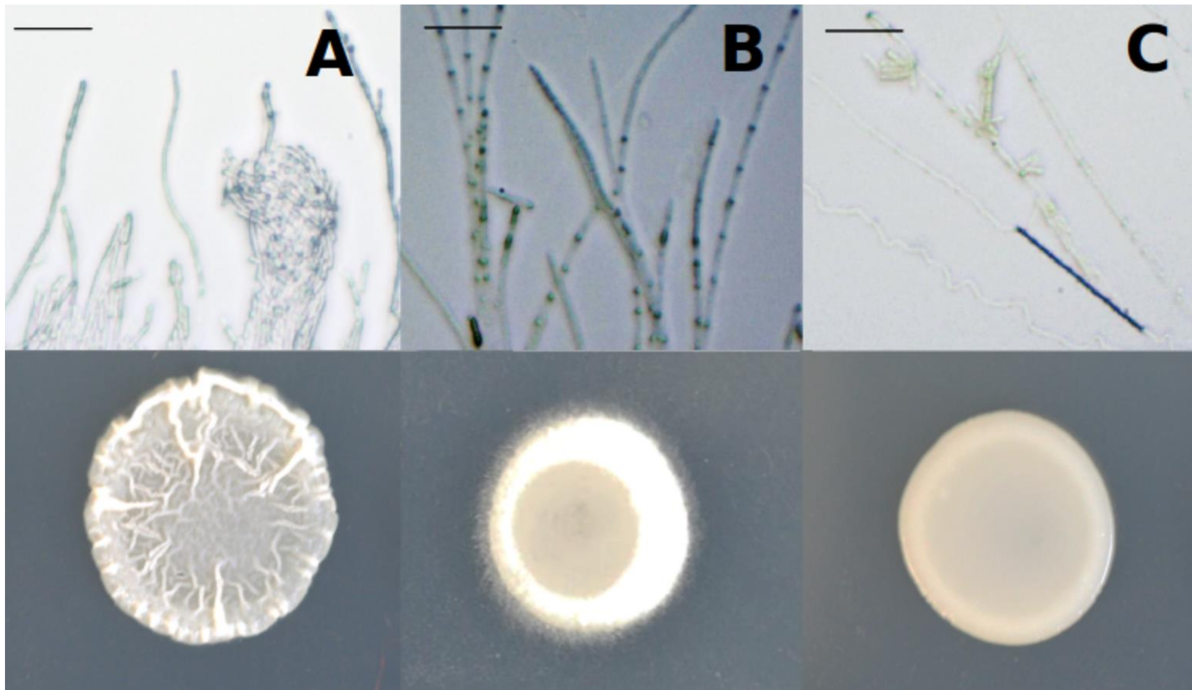

Figure 6. Interspecific Hybridization assay between *S. panici-leucophaei* and *S. scitaminuem*. (A) Positive control (*S. panici-leucophaei*). (B) Positive control (*S. scitamineum*) (C) Positive interspecific hybridization (Spl *vs* Ssci). Scale bars represent 20uM.

Sporidia of opposite sexual types (A and B) emit conjugation (reproductive) hyphae, thus enabling the occurrence of sexual reaction leading to anastomosis (hyphae fusion). The sexual reaction between compatible sporidia occurs through a pheromone and membrane receptors recognition system controlled by two loci, a and b, a process conserved among smuts. Even though *S. panici-leucophaei* and *S. scitamineum* are different smut species, their mating-type system is compatible, probably due to their capability of recognizing interspecific pheromone and membrane receptors.

3.4.) ITS-base phylogeny

The phylogeny of ten different smut species was constructed using Internal Transcribed Spacer (ITS) sequences recovered from GenBank accessions. The ITS of *S. panici-leucophaei* was isolated and sequenced, generating a 565 bp sequence that was included for the phylogeny reconstruction. The ITS - based tree showed separation among smut genus *Sporisorium* and

Ustilago (Fig. 7). These observations are limited based on the number of samples used, but maybe relevant, considering the evolution of these pathogens, and should be further investigated.

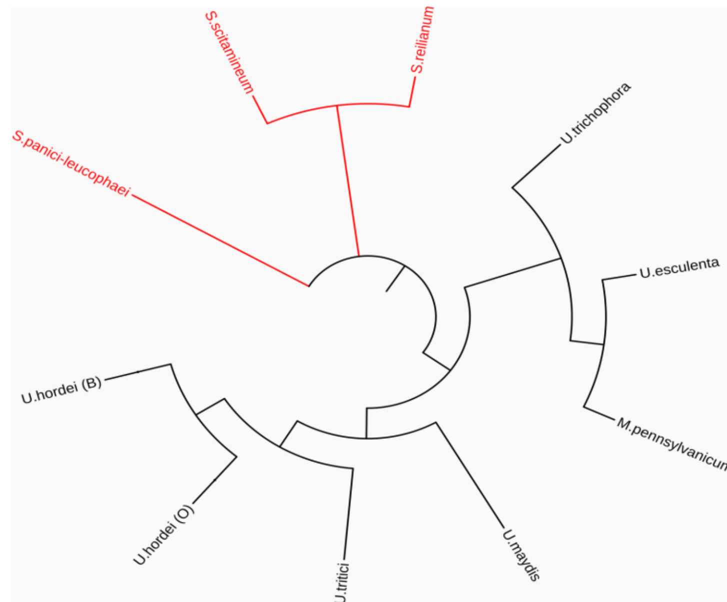

Figure 7. Phylogenetic tree among smuts. Internal Transcribed Spacer (ITS)-based tree using GenBank accessions of ten different smuts - *maximum likelihood* method, using the MEGA5 software - bootstrap test (10,000 repetitions). Red branches indicate fungus from the *Sporisorium* genus.

3.5.) Genome Assembly

Illumina sequencing allowed us to assemble a first draft genome of *S. panici-leucophaei* Spl10A. A total of 715.363 reads representing 430 Mb of raw data were used to generate an assembly that comprised approximately 18,80 Mbp distributed in 561 contigs with an average coverage of 10x (Table 2). For this first assembly (v1), we used *S. scitamineum* complete sequence as a reference. We then executed Canu (v1.8) (with parameter correctedErrorRate=0.15) using only nanopore reads that aligned to the Illumina assembly (v1). This strategy was necessary because approximately 49% of all nanopore reads were from a bacterial contaminant of the cell culture, identified using BlastN against the NCBI database (NCBI, 2020). Therefore, from a total of 830 Mbp, 124,834 reads (423 Mbp) were used to generate a new assembly (v2) with 50 contigs and an average coverage of 21x (Table 2). Reads smaller than 1,000 bp were not used at this stage. Indeed, the contaminant reads assembled belonged to the genome of a contaminant bacteria. Other than that, no other contig was part of the *S. panici-leucophaei* genome except that containing the mitochondrial DNA sequence.

TABLE 2 | Sequencing output and genome assembly statistics.

	Illumina	Nanopore (MinION)
Raw data (Mbp)	430*	830*
Assembly program	SPADES	Canu
Coverage	10x	21x
Genome size (Mbp)	18.8	18.9
Contigs	561	22
N50	59,88	968,948

* Contamination was detected within sequenced reads

We then implemented Pilon (v1.18) to correct indels using Illumina reads, and Cross_match to close the gaps to generate the complete sequence of *S. panici-leucophaei* genome comprising 18,913,471 nucleotides distributed in 22 contigs (v3, chromosomes with no gaps within them) (Table 3). Chromosome 01 was considered the largest one consisted of 2,101,984 bp, while, chromosome 22 presented less than 155,000 bp.

TABLE 3 | *S. panici-leucophaei* genome assembly v3 coverage *per* chromosome.

Chromosome	Length (bp)	Read counts (Illumina)	Average coverage (Illumina)
01	2,101,984	157,169	21,44
02	1,766,037	97,939	15,93
03	1,576,904	77,788	14,17
04	1,477,939	77,472	15,03
05	1,085,824	60,625	16,04
06	985,986	55,695	16,21
07	968,948	52,135	15,43
08	925,274	46,490	14,41
09	913,032	49,163	15,47
10	797,769	42,297	15,19
11	717,528	32,977	13,13
12	648,462	31,923	14,08
13	631,431	32,380	14,66
14	579,986	30,180	14,93
15	564,423	31,625	16,07
16	546,752	25,299	13,21
17	529,375	26,016	14,05
18	515,699	28,418	15,76
19	501,296	26,810	15,34
20	449,585	22,732	14,51
21	398,396	20,192	14,48
22	152,790	8,074	15,21

Next, we used BWA to align the Illumina reads to the v3 and exhibit the Illumina reads coverage on the final assembly (Table 3). On average, we covered each chromosome with a 15-fold of Illumina reads. Chromosome 01 has largest coverage (~22-fold) because it encloses the rDNA genes. In S. *scitamineum*, Taniguti *et al*. (2015) estimated 130 copies of rDNA genes are

in chromosome 02. Here we were able to assemble four copies of the complete rDNA sequence, including the 18S, 5.8S, and 28S genes, and respective ITS and IGS. Telomeric motifs were identified for all 22 chromosomes.

3.6.) Genome and Protein Annotations

The *S. panici-leucophaei* Spl10A genome included 6,404 protein-encoding genes (CDS) (Table 4). Using a Basidiomycota database as reference (with 1335 BUSCOs groups), BUSCO *v3* mode genome (Waterhouse *et al*., 2017) output suggested a 97.0% completeness for *S. panici-leucophaei* v3 assembly (see Suppl. Table 2). Gene density (Genes/ 10 Kbp) was similar among all chromosomes (Table 4) with an average of 3.3 genes *per* 10,000 bp. The chromosome with least number of genes is 22 with 2.3 genes every 10 Kbp. The mtDNA sequence present a high gene density with 9.4 genes every 10,000 bp.

TABLE 4 | Predicted gene density for *S. panici-leucophaei* genome assembly v3.

Chromosome	Length (10 Kbp)	Genes	Genes / 10 Kbp
Predicted genes for *S. panici-leucophaei* gDNA			
	6,404		
Predicted genes *per* chromosome			
Chromosome	**Length (10 Kbp)**	**Genes**	**Genes / 10 Kbp**
01	210	719	3.4
02	176	588	3.3
03	157	615	3.9
04	147	505	3.4
05	108	381	3.5
06	98	315	3.2
07	96	332	3.4
08	92	306	3.3
09	91	281	3.0
10	79	258	3.2
11	71	269	3.7
12	64	217	3.3
13	63	212	3.3
14	57	198	3.4
15	56	177	3.1
16	54	203	3.7
17	52	170	3.2
18	51	175	3.4
19	50	166	3.3
20	44	160	3.6
21	39	122	3.1
22	15	35	2.3
Predicted genes for *S. panici-leucophaei* mtDNA			
Chromosome	**Length (10 Kbp)**	**Genes**	**Genes / 10 Kbp**
Circular	7.8	73	9.4

Repetitive elements represented 2.70 % (511,054 bp) of the *S. panici-leucophaei* genome (Table 5). Most of the repeated elements were identified simple repeats or low complexity elements. Transposable elements were of two types, retroelements of SINE family and DNA transposons of the Tc1-IS630-Pogo, commonly find in fungal genomes (Muszewska *et al*., 2011; Soliai *et al*., 2014).

TABLE 5 | Identified interspersed and simple repeats in *S. panici-leucophaei* genome (RepeatMask)

	Number of elements	Elements sizes (bp)	Percentage (%)
Retroelements			
Sine	28	3,798	0.02
DNA transposons			
Tc1-IS630-Pogo	39	32,833	0.17
Unclassified	308	115,171	0.80
Total interspersed repeats	--	151,802	0.80
Small RNA	6	11,695	0.06
Satellites	8	11,710	0.01
Simple repeats	8292	312,983	1.65
Low complexity	681	33,403	0.18

The mitochondrial genome is composed of 78,051 bp and presented the standard 15 protein-encoding genes (nad1, nad2, nad3, nad4, nad4L, nad5, nad6, cox1, cox2, cox3, cob, atp2, atp6, atp8, and rps3) (see Suppl. Table 3) along with 34 unknown hypothetical ORFs. Among the RNA encoding genes, 23 tRNA genes were detected representing all 20 amino acids, and two tRNAs for each Ser, Met and Arg. The last nucleotide of the nad2 gene is the first one of the nad3 gene. Among all 73 genes detected in the mtDNA (including copies), only 3 presented introns: cob, cox1 and nad5. The cox1 gene was the most intron rich gene with 7 introns, followed by cob and nad5 with 4 and 2, respectively. Moreover, interestingly, hypothetical ORF156 presented a second gene copy with equal sequence length (470 bp).

3.7.) Comparative Genomics

Smut (Ustilaginomycotina) genomes available in public databases and the v3 assembly of *S. panici-leucophaei* were used as input for OrthoMCL software (Table 6). A total of 7,277 ortho-groups were described among the ten smut species analysed. For *S. panici-leucophaei*, from a total 6,404 genes, 5,955 were considered single-copy and 449 had paralogs. 6,140 genes were assigned to a certain ortho-group, whereas, 264 genes were considered singletons (genes not assigned to any OrthoMCL cluster). Sourgrass smut presented 5 unique groups (exclusive groups) containing 11 genes. *S. scitamineum* demonstrated a similar pattern, with 6,127 single-copy genes and 550 paralogs from a total 6,677 gene repertoire. Within these genes, 6,319 were assigned to a certain cluster, while, 358 genes were considered singletons. *S. scitamineum* presented 14 unique groups containing 49 genes.

From a total 7,891 genes, 6,657 were single-copy ones for *U. hordei* (O), the smut with the highest number. Curiously, *U. hordei* (O) presented 1235 paralog genes, while *U. esculenta* presented 876 genes.

TABLE 6 | Genomic statistic of smut fungi.

Genomic statistics	***Spl***	***Ssci***	***Srei***	***UhoO***	***UhoB***	***Umay***	***Uesc***	***Utri***	***Utcp***	***Mpen***
Assembly										
Total assembly size (Mbp)	18.91	19.95	18.38	24.63	21.15	19.64	20.19	18.63	20.68	19.23
Average base coverage	21x	500x	29x	487x	25x	10x	139x	278x	na	339x
Number of contigs (>500 bp)	22	26	45	2200	713	27	298	73	215	435
N50 (bp)	968,948	875,830	772,363	39,442	307,727	884,984	403,507	610,801	179,640	121,670
Largest contig (bp)	2,101,984	2,009,762	2,448,206	171,399	542,606	2,476,501	1,882,320	1,118,949	637,988	690,500
GC-content (%)	53.99	55.16	59.87	51.60	52.16	54.03	54.42	57.08	53.03	50.90
Coding sequences										
Number of genes	6,404	6,677	6,675	7,891	7,111	6,784	6,773	6,776	6,499	6,280
SIngle-copy genes	5,955	6,127	6,223	6,657	6,528	6,225	5,897	6,123	5,914	5,832
Co-orthologs groups	5,892	5,970	6,198	6,334	6,350	6,084	5,774	6,054	5,843	5,557
Genes in the group	6,140	6,319	6,441	7,196	6,701	6,403	6,332	6,407	6,196	5,800
Paralogs	449	550	452	1234	583	559	876	653	585	448
Unique groups	5	14	5	98	20	13	35	2	15	8
Singletons	264	358	234	696	410	381	441	369	303	480
Repeat sequences (%)										
Total of bases masked	2.70	5.85	2.68	25.12	13.21	4.13	9.65	2.34	4.18	3.88

3.8.) Comparative Secretome and Effectorome

Secretome prediction in *S. panici-leucophaei* and *S. scitamineum* predicted proteomes was accomplished applying a pipeline consisted of: scanning proteins that present a signal peptide, selecting those that does not present a transmembrane domain and an endoplasmic reticulum (ER) retention signal. As discriminated at Table 7, from a total of 6,404 predicted proteins, only 382 were predicted as secreted for *S. panici-leucophaei*. For *S. scitaminuem*, 388 proteins were considered secreted from a total 6,677 protein repertoire.

TABLE 7 | Predicted proteome, secretome, effectors and cazymes for *S. panici-leucophaei* and *S.scitamineum*.

Genome characteristics	*S. panici-leucophaei*	*S. scitamineum*
Predicted Proteome	6,404	6,677
Predicted Secretome	382	388
Predicted Effectors	66	68
Predicted Cazymes	343	273

Proteins potentially associated with degradation, modification, or creation of glycosidic bonds, denominated CAZymes were also predicted for both smuts. For *S. panici-leucophaei,* 343 CAZymes were identified, whereas, for *S. scitamineum*, 273 proteins were predicted.

From the predicted secretome, we could screen for effector candidates for both smuts, using EffectorP. Both pathogens presented a similar effector repertoire where approximately 17% of all secreted proteins were considered effector candidates. For the sugarcane smut fungus, 68 proteins were considered effectors, whereas, 66 proteins were denominated effectors for *S. panici-leucophaei*. Among them, 25 proteins were shared between both fungi, while, 41 were unique to *S. panici-leucophaei* and 43 were unique to *S. scitamineum* (Fig. 8).

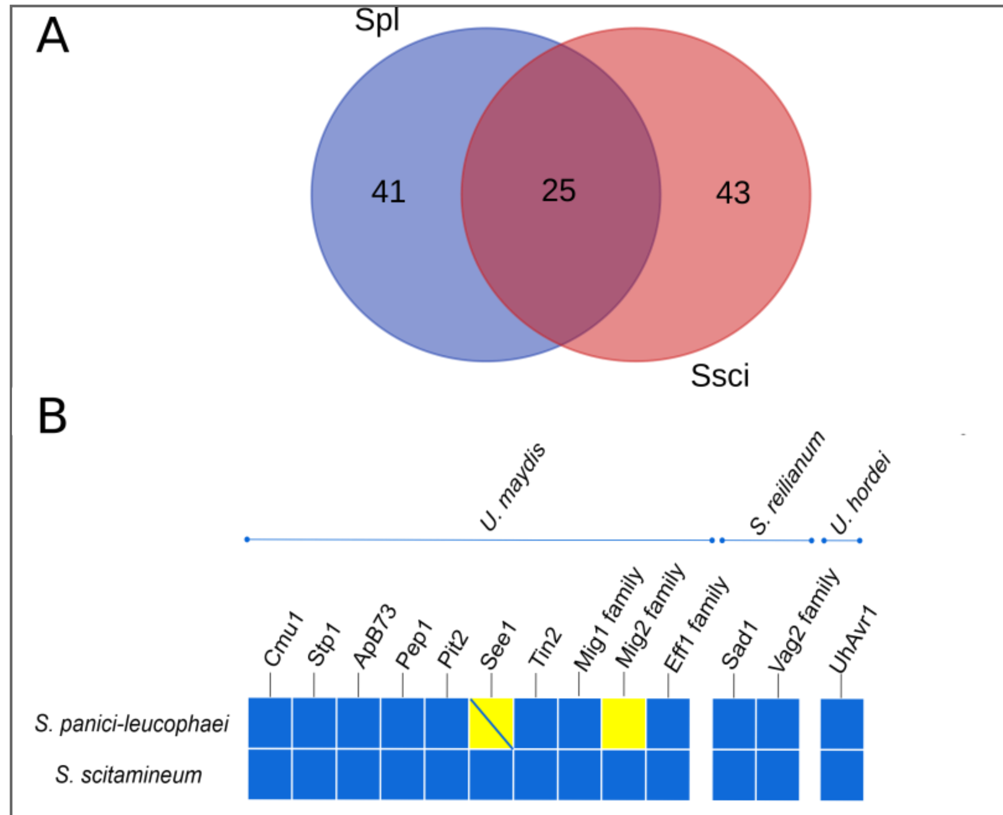

Figure 8. Candidate effector prediction in *S. panici-leucophaei* and *S. scitamineum*. (A) Venn diagram of showed candidate effector prediction for both smuts. Colors represent the exclusive candidates from both smuts and the intersection represents shared candidates. (B) Presence (blue squares) and absence (yellow squares) of genes or gene families encoding known effectors in *U. maydis*, *S. reilianum* and *U. hordei* based on OrthoMCL. Crossed yellow squares indicate that a similar genic region is present in the genome by tblastn.

Characterized effectors in *U. maydis*, *S. reilianum* and *U. hordei* (B) were screened for orthologous in *S. panici-leucophaei* and *S. scitamineum* using the output from OrthMCL. Effectors present in both predicted proteomes were identified, including Cmu1 (*Chorismate mutase 1*), Stp1 (*Stop after penetration 1*), ApB73 (*Apathogenic in B73*), Pep1 (*Protein essential during penetration-1*), Pit2 (*Protein involved in tumors 2*), Tin2 (*Tumor inducing 2*), members of the Mig1 (*Maize-induced gene 1*) and Eff1 family, Sad1 (*Suppressor of apical dominance 1*), UhAvr1 and members of the Vag2 family (Fig. 8). Initially, orthologous of two characterized effectors, See1 (*Seedling efficient effector1*) and members of the Mig2 family, were found only for *S. scitamineum*. However, tBlastN analysis revealed the presence of See1 in *S. panici-leucophei.*

4.) Discussion

4.1.) Experimental approaches to study *S. panici-leucophaei* biology

This work presented the first initiative to improve the understanding of similarities between the disease signs produced by *S. panici-leucophaei* and *S. scitamineum*. Both smut pathogens modulate host physiology to induce fungal sporogenesis within a structure called whip. We implemented experimental approaches to study the *S. panici-leucophaei* genome.

The time-points established for germination, appressorium formation, and colonization of host tissues were 8 hpi, 48 hpi, and 72 hpi, respectively, which are similar to the responses detected for sugarcane plants to *S. scitamineum* (Peters *et al*., 2017). The inoculated plants maintained in a greenhouse for more than 120 days did not produce whip (see Supple. Fig. 1). There is no information about the status of *D. insularis* resistance toward smut infection. The genotype used in our study was not the same as the one we first isolated teliospores. However, our first goal was to determine steps to allow the recognition of S. *panici-leucophaei* structures, which showed to be similar to those of other smut fungi (Zuo *et al*., 2019). Recently, the CSIRO (Australia's National Science Agency) released a smut fungi biocontrol for an environmental weed *Tradescantia fluminensis* (an invasive weed from South America) threatening native vegetation and clogging waterways in Australia (CSIRO, 2019). Because sourgrass is a weed resistant to glyphosate, our work may also provide relevant information for this line of research.

The study of interspecific hybridization in fungi helps the understanding of the evolution and emergence of new pathogens (Büker *et al*., 2013; Stukenbrock, 2016). Among the best examples of well-studied systems for interspecific hybridization are those of the anther smut pathogens *Microbotryum lychnidisdioicae* and *M. silenes-dioicae*, regularly found in nature (Depotter *et al*., 2016).

D. insularis is a weed found invading various crop fields, including sugarcane. We identified compatible mates of *S. panici-leucophaei* and *S. scitamineum*. The genomic analysis confirmed the results obtained for the fusion of compatible sporidia visualized in our experiments. The protein sequence of the pheromone type-A (pra2) of *S. scitamineum* SSC04 is 86% similar to predicted protein pra2 encoded by the genome of *S. panici-leucophaei* (data not shown). The protein pra1 encoded by the available SSC39 genome is associated with the mating-type B (Monteiro-Vitorello *et al*., 2018). The impact of this hybrid, if any, considering viability and host colonization still needs further analysis.

4.2.) An overview of *S. panici-leucophaei* genome features

We defined 22 chromosomes of *S. panici-leucophaei* with no gaps using a combination of Oxford Nanopore and Illumina technologies. Our work is the first report of a smut-fungus genome sequenced entirely using the new MinION platform and the second available with all

chromosomes sequenced from telomere-to-telomere (Taniguti *et al*., 2015). These data have the potential to initiate a line of investigation considering the genome structure of smut-fungi related to their evolution. An approach considering the evolution of the mating-type loci and the variation of bipolar and tetrapolar sexuality patterns is associated with the disruption through genome reorganization of the two mating loci (*a* and *b*) (Bakkeren *et al*., 2008; Taniguti *et al*., 2015). For instance, the mating-type loci of *S. scitamineum* segregate as one genetic unity of the chromosome 2 (bipolar pattern), the mating-type loci of *U. maydis* located in two distinct chromosomes (tetrapolar pattern) (Bakkeren *et al*., 2008).

Despite the elevated number of contigs in the assembly, *U. hordei* (O) has the largest genome size among smuts with 24.63 Mbp, followed by the other *Ustilago hordei* (B) species with 21.15 Mbp. Five species have smaller genome sizes than 20 Mbp: *S. scitaminuem*; *M. pennsylvanicum*; *S. panici-leucophaei*; *U. tritici* and *S. reilianum* with the latter being the smallest one with 18.38 Mbp. The percentage of GC-content ranges from 50.90 in *M. pennsylvanicum* to 59.87 in *S. reilianum*. *S. panici-leucophaei* present a genome size closer to those of *S. scitamineum* and *S. reilianum* with an average GC content of 54%. Repetitive sequences, represented as the percentage of masked base pairs, demonstrate large amplitude within smut pathogens. *U. tritici* presented 2.34% of bases masked, while *U. hordei* (O) presented the largest amount with 25.12%, more than a quarter the size of its genome assembly.

Smut genomes contain relatively small amounts of repetitive elements as compared to other fungi (0.02 to 29.8%), usually not higher than 10% (Benevenuto *et al*., 2018; Castanera *et al*., 2016). *U. hordei*, among smuts, is an exception (25%). *S. panici-leucophaei* is among the smuts with the least number of repetitive elements (2.7%). A mechanism of RIP (Repeat-Induced Point Mutation) usually controls the proliferation of transposable elements in basidiomycetes (Horns *et al*., 2012). Remains to be investigated in the recently sequenced smut genomes the functioning of such mechanism.

4.3.) Overview of the predicted protein pool of *S. panici-leucophaei* compared with *S. scitamineum*

The total number of 6,404 proteins predicted for *S. panici-leucophaei* lays within the average of the number of genes described for the Ustilaginaceae genomes sequences to date (6797 ± 499). The gene repertoire varied from 6,280 predicted genes in the wild plant pathogen *M. pennsylvanicum* to 7,892 predicted genes in *U. hordei* (O). Three smuts presented similar gene sets when compared to *S. leucophaei-panici*: both *S. reilianum* and *U. tritici* with 6,776

predicted genes followed by *S. scitamineum* with 6,677 gene predictions. Benevenuto *et al.* (2018) showed that indeed, smuts contain a very conserved pool of proteins, making the understanding of host specificity a challenging task. The authors suggested that the pool of candidate effectors may be determinants of host specificity.

S. panici-leucophaei genome encodes proteins that lay within most of the orthogroups formed when the proteins encoded by smuts are analyzed altogether, reinforcing the conclusions of Benevenuto *et al*. (2018). Comparing *S. panici-leucophaei* and *S. scitamineum,* the numbers of proteins in each group, the Cazymes, and those detected as unique revealed that potentially the strategies used by both fungi to invade and promote disease symptoms in their hosts would be equivalent. Therefore, the use of *S. panici-leucophaei* and *D. digitalis* as a model to understand smut disease and the induction of whip as the main sign of the disease is very attractive.

4.4.) Effectorome of *S. panici-leucophaei*

Only two candidate effectors were detected as specific of *S. panici-leucophaei* compared to those found in *S. scitamineum* in our data using EffectorP (Fig 8B). However, a closer analysis revealed that at least one effector is potentially encoded by the *S. panici-leucophaei*. We identified an ortholog of the *see1* gene by using searches (tblastn) in the complete genome sequence. The other missing effector belongs to the Mig2 family of effectors (Benevenuto *et al*., 2018).

See1, for instance, is a common effector among smut species (Benevenuto *et al*., 2018), functionally characterized in infections caused by *U. maydis* and *U. hordei* (Redkar *et al*., 2015a; b). In the interaction of *U. maydis* in maize, the See1 effector is required for tumor progression in leaves. Whereas, in infections caused by *U. hordei* in barley, the See1 protein seems to not be functional (Redkar *et al*., 2015b). *U. hordei* does not induce tumors in leaves of barley. So far, reports describe that only eventually, *S. scitamineum* induces tumors in leaves of sugarcane, and we have no information about *S. panici-leucophaei*. Overall, See1 is highly diversified, presenting only one conserved block in the central region of the protein that has above 50% identity (Redkar *et al*., 2015).

Migs are a family of effectors first identified in *U. maydis* (Basse *et al*., 2002; Farfsing *et al*., 2007). Genes encoding Migs are organized in a cluster in the *U.maydis* genome. These genes express equally four days after inoculation, whereas expression is absent or only faintly detectable during axenic growth (Basse *et al*., 2002). The specific role of Migs was not yet

described, but their expression is necessary for full virulence. *S. scitamineum* does not present a full set of Mig genes, and those were potentially lost in *S. panici-leucophaei*.

These types of differences in function may be another relevant factor to understand smut diseases. Despite the similar set of proteins encoded by them, the specific function, particularly of the effectors, may vary among them.

A similar set of genes and similar disease symptoms, in this case, reinforces the relevance of the use of S. *panici-leucophaei* and *D. digitalis* as a model, particularly considering pathogens of sugarcane. Sugarcane has a highly complex genome resulting from the interspecific hybridizations of breeding programs, besides plant architecture and long reproductive development not suitable for lab experiments (Thirugnanasambandam *et al*., 2018).

5.) Conclusion

With this work, we provided experimental data and genomic information about the smut fungus *S. panici-leucophaei*. We could characterize the smut life cycle including the quantification of specific fungal structures, related to host penetration and *in planta* colonization. The interspecific relationship among smuts was evaluated by a hybridization assay, where, *S. panici-leucophaei* and *S. scitamineum* produced conjugation hyphae, suggesting compatibility among them. In addition, we completely sequenced the sourgrass smut genome using a hybrid strategy with Illumina and Oxford Nanopore reads, generating an assembly composed of 22 chromosomes and a circular mtDNA sequence. We identified the effectorome repertoire for *S. panici-leucophaei* and compared with *S. scitamineum,* highlighting specific and shared genes that have a potential role in host–pathogen interaction. With this current project, we also brought new knowledge to a poorly studied smut, shedding light on novel opportunities for this particular and complex pathosystem.

REFERENCES

ALEXANDER K.C., RAMAKRISHNAN K. (1980). Infection of the bud, establishment in the host and production of whips in sugarcane smut (*Ustilago scitaminea*) of sugarcane. Proc Int Soc Sugar Cane Technol, p.1452–1455.

ALEXANDER K., SRINIVASAN K. (1966). Sexuality in Ustilagoscitaminea Syd. Current Science 35 : 603–604.

APOGA D., BARNARD J., CRAIGHEAD H.G., HOCH H.C. (2004). Quantification of substratum contact required for initiation of *Colletotrichum graminicola appressoria*. Fungal Genet Biol 41:1–12 762 Planta (2017) 245:749–764.

ARMENTEROS J.J.A., TSIRIGOS K.D., SØNDERBY C.K., PETERSEN T.N., WINTHER O., BRUNAK S., *et al*. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature biotechnology, 37(4), 420.

BASSE C.W., KOLB S., KAHMANN R., (2002). A maize-specifically expressed gene cluster in *Ustilago maydis*. Molecular microbiology, 43(1), 75-93.

BAUCH R., (1923). Über'Ustilago longissima'und ihre Varietät Macrospora. Verlag von G. Fischer.

BEGEROW D., GÖKER M., LUTZ M., STOLL M. (2004). On the evolution of smut fungi on their hosts. Frontiers in basidiomycote mycology, p. 81–98.

BEGEROW D., SCHÄFFER A.M., KELLNER R., YURKOV A., KEMLER M., OBERWINKLER F. (2014). Ustilaginomycotina. In: MCLAUGHLIN, D.; SPATAFORA, J. W. (Eds.).The Mycota. vol. 7. Part A. Systematics and evolution. 2. ed. Berlin: Springer, p. 295–329.

BAKKEREN G., KÄMPER J., SCHIRAWSKI J. (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. Fungal Genetics and Biology, v. 45, p. S15–S21.

BENDTSEN J.D., JENSEN L.J., BLOM N., VON HEIJNE G., BRUNAK S. (2004). Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. 17, 349–356.

BENEVENUTO J., LONGATTO D.P., REIS G.V., MIELNICHUK N., PALHARES A.C., CARVALHO G., *et al.* (2016). Molecular variability and genetic relationship among Brazilian strains of the sugarcane smut fungus. FEMS Microbiology Letters, v. 363, n. 24, p. fnw277.

BENEVENUTO J., TEIXEIRA-SILVA N.S., KURAMAE E.E., CROLL D., MONTEIRO-VITORELLO C.B. (2018). Comparative Genomics of Smut Pathogens: Insights From Orphans and Positively Selected Genes Into Host Specialization. Frontiers in Microbiology, v. 9, p. 660.

BÖLKER M., URBAN M., KAHMANN R., (1992). The a mating type locus of *U. maydis* specifies cell signaling components. Cell, 68(3), 441-450.

BREFORT T., TANAKA S., NEIDIG N., DOEHLEMANN G., VINCON V., KAHMANN, R. (2014). Characterization of the largest effector gene cluster of *Ustilago maydis*. PLoS pathogens, 10(7), e1003866.

BÜKER B., PETIT E., BEGEROW D., HOOD M.E., (2013). Experimental hybridization and backcrossing reveal forces of reproductive isolation in Microbotryum. BMC evolutionary biology, 13(1), 224.

CARVALHO G., QUECINE M.C., LONGATTO D.P., PETERS L.P., ALMEIDA J.R., SHYTON T.G., *et al*. (2016). *Sporisorium scitamineum* colonisation of sugarcane genotypes susceptible and resistant to smut revealed by GFP-tagged strains. Annals of Applied Biology, 169(3), 329-341.

CASTANERA R., LOPEZ-VARAS L., BORGOGNONE A., LABUTTI K., LAPIDUS A., SCHMUTZ J., *et al*. (2016). Transposable elements versus the fungal genome: impact on whole- genome architecture and transcriptional profiles. PLoS genetics, 12(6), e1006108.

CHRISTENSEN J.J. (1963). Corn smut caused by Ustilago maydis. The American Phytopathological Society, p. 1–41.

COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION (CSIRO). (2019). https://www.csiro.au/en/News/News-releases/2019/Super-smut-fungus-to-combat- smothering-weed.

CONESA A., GÖTZ S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. International journal of plant genomics, v. 2008, p. 619832.

DALIO R.J.D., MAGALHÃES D.M., ATÍLIO L.B., RODRIGUES C.M., BRETON M.C., PICHI S., *et al*. (2014). Efetores na interação planta-patógeno. Revisão Anual de Patologia de Plantas, Passo Fundo, v. 22, p. 25-68.

DE CASTRO E., SIGRIST C.J., GATTIKER A., BULLIARD V., LANGENDIJK-GENEVAUX P.S., GASTEIGER E., *et al*. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic acids research, 34(suppl_2), W362-W365.

DEPOTTER J.R., SEIDL M.F., WOOD T.A., THOMMA B.P., (2016). Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. Current Opinion in Microbiology, 32, 7-13.

DIAZ-QUIJANO C., WICHMANNA F., SCHLAICHA T., FAMMARTINOA A., HUCKAUFB J., SCHMIDTC K., *et al*. (2016). KP4 to control *Ustilago tritici* in wheat: Enhanced greenhouse resistance to loose smut and changes in transcript abundance of pathogen related genes in infected KP4 plants. Biotechnology Reports, v. 11, p. 90–98.

DJAMEI A., SCHIPPER K., RABE F., GHOSH A., VINCON V., KAHNT J. et al., (2011). Metabolic priming by a secreted fungal effector. Nature, 478(7369), P.395.

DOEHLEMANN, G., REISSMANN, S., ASSMANN, D., FLECKENSTEIN, M. AND KAHMANN, R. (2011). Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis* induced-tumour formation. *Molecular Microbiology*, *81*(3), pp.751-766.

EDGAR R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

FARFSING J.W., AUFFARTH K., BASSE C.W., (2005). Identification of cis-active elements in *Ustilago maydis* mig2 promoters conferring high-level activity during pathogenic growth in maize. Molecular plant-microbe interactions, 18(1), 75-87.

HAUEISEN J., STUKENBROCK E.H. (2016). Life cycle specialization of filamentous pathogens - colonization and reproduction in plant tissues. Current Opinion in Microbiology, v. 32, p. 31–37.

HEMETSBERGER C., HERRBERGER C., ZECHMANN B., HILLMER M., DOEHLEMANN G. (2012). The Ustilago maydis effector Pep1 suppresses plant immunity by inhibition of host peroxidase activity. PLoS Pathogens, 8(5), e1002684.

HOFF K.J., STANKE M., (2019). Predicting genes in single genomes with augustus. Current protocols in bioinformatics, 65(1), e57.

HORST R.K. (2013). Smuts. In: Westcott's Plant Disease Handbook. Dordrecht: Springer Netherlands, p. 373–379.

HORNS F., PETIT E., YOCKTENG R., HOOD M.E., (2012). Patterns of repeat-induced point mutation in transposable elements of basidiomycete fungi. Genome biology and evolution, 4(3), 240-247.

HORRIDGE G.A., TAMM S.L.,(1969). Critical point drying for scanning electron microscopy study of ciliar motion. Science 13: 871-818.

JAMES G.L., (1973). Smut spore germination on sugarcane internode surfaces. Proceedings of The South African Sugar Technolo-gists' Association, pp 179–180

KÄLL L., KROGH A., SONNHAMMER E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027–1036.

KAMPER J., KAHMANN R., BOLKER M., MA L.-J., BREFORT T., SAVILLE B.J., *et al.*, (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*.Nature 444, 97–101

KELLNER R., *et al*. (2011). Interspecific sex in grass smuts and the genetic diversity of their pheromone-receptor system. PLoS genetics, v. 7, n. 12, p. e1002436.

KIRZINGER M.W.B., STAVRINIDES J. (2012). Host specificity determinants as a genetic continuum. Trends in Microbiology, v. 20, n. 2, p. 88–93.

KISSMANN K.G. (1997). Plantas infestantes e nocivas. São Paulo: Basf Brasileira.

KOREN S., WALENZ B.P., BERLIN K., MILLER J.R., PHILLIPPY A.M., (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Research.

KROGH A., LARSSON B., VON HEIJNE G., SONNHAMMER E.L., (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology, 305(3), p. 567-580.

LANVER D., TOLLOT M., SCHWEIZER G., LO PRESTI L., REISSMANN S., MA L. S. (2017). Ustilago maydis effectors and their impact on virulence. Nat. Rev. Microbiol. 15, 409–421.

LAURIE J.D., ALI S., LINNING R., MANNHAUPT G., WONG P., GÜLDENER U., *et al.*, (2012). Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. Plant Cell 24, p. 1733–1745

LI L., STOECKERT C.J., ROOS D.S., (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes.Genome Res. 13, 2178–2189. doi: 10.1101/gr. 1224503

LIANG L. (2013). The role of Stp1, a secreted effector, in the biotrophic interaction of *Ustilago maydis* and its host plant maize (Doctoral dissertation, Philipps-Universität Marburg).

LOMBARD V., GOLACONDA RAMULU H., DRULA E., COUTINHO P.M., HENRISSAT B., (2013).The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids research, 42(D1), D490-D495.

LU H., GIORDANO F., NING Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. Genomics, proteomics & bioinformatics, v. 14, n. 5, p. 265-279.

MACHADO A.F.L., FERREIRA L.R., FERREIRA F.A., FIALHO C.M.T., TUFFI SANTOS L.D., MACHADO M. (2006). Análise de Crescimento de *Digitaria insularis*. Planta Daninha, p. 641–647.

MACHADO A.F.L, MEIRA R.M.S, FERREIRA L.R., FERREIRA F.A., FIALHO C.M.T., TUFFI SANTOS L.D., *et al*. (2008). Caracterização Anatômica de Folha, Colmo e Rizoma de *Digitaria insularis*. Planta Daninha, v. 26, n. 1, p. 1–8.

MONDO V.H.V., CARVALHO S.J.P. de, DIAS A.C.R., MARCOS FILHO J. (2010). Efeitos da luz e temperatura na germinação de sementes de quatro espécies de plantas daninhas do gênero *Digitaria*. Revista Brasileira de Sementes, v. 32, n. 1, p. 131–137.

MONTEIRO-VITORELLO, C. B., SCHAKER, P. D. C., BENEVENUTO, J., TEIREIRA-SILVA, N., ALMEIDA, S. (2018). "Progress in understanding fungal diseases affecting sugarcane: smut," in Achieving sustainable cultivation of sugarcane, ed. P. Rott (Burleigh Dodds Science Publishing).

MUSZEWSKA A., HOFFMAN-SOMMER M., GRYNBERG M., (2011). LTR retrotransposons in fungi. PLoS One, v. 6, n. 12, p. e29425.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2020 March 03]. Available from: https://www.ncbi.nlm.nih.gov/.

PETERS L.P., CARVALHO G., VILHENA M.B., CRESTE S., AZEVEDO R.A., MONTEIRO-VITORELLO C.B. (2017). Functional analysis of oxidative burst in sugarcane smut-resistant and -susceptible genotypes. Planta, v. 245, n. 4, p. 749–764.

PIĄTEK M. (2006). *Sporisorium kenyanum*, a new smut fungus with long twisted sori on *Setaria pallide-fusca* in Kenya. Polish Botanical Journal, v. 51, n. 2, p. 159–164.

POREBSKI S., BAILEY L.G., BAUM B.R., (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter, 15(1), p. 8-15.

REDKAR A., HOSER R., SCHILLING L., ZECHMANN B., KRZYMOWSKA M., WALBOT V., *et al*. (2015). A secreted effector protein of *Ustilago maydis* guides maize leaf cells to form tumors. The Plant Cell, 27(4), p. 1332-1351.

REDKAR A., VILLAJUANA-BONEQUI M., DOEHLEMANN G., (2015). Conservation of the *Ustilago maydis* effector See1 in related smuts. Plant signaling & behavior, 10(12), e1086855.

ROLAND A. M. 1978. General preparations and staining of thin sections. In: Hall JL ed.Electron microscopy and cytochemistry of plant cells .NewYork: Elsevier, 1–62.

SCHATTNER P., BROOKS A.N., LOWE T.M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research, v. 33, n. Web Server, p. W686–W689.

SCHIRAWSKI J., MANNHAUPT G., MÜNCH K., BREFORT T., SCHIPPER K., DOEHLEMANN G., *et al*. (2010). Pathogenicity determinants in smut fungi revealed by genome comparison. Science 330, p. 1546–1548.

SELIN C., KIEVIT T.R., BEL MONTE M.F., FERNANDO D.W.G. (2016). Elucidating the role of effectors in plant-fungal interactions: progress and challenges. Frontiers in Microbiology 7: 600.

SHARMA R., MISHRA B., RUNGE F., THINES M., (2014). Gene loss rather than gene gain is associated with a host jump from monocots to dicots in the Smut Fungus *Melanopsichiumpennsylvanicum*.Genome Biol. Evol. 6, p. 2034–2049.

SINGH N., SOMAI B., PILLAY D. (2004). Smut disease assessment by PCR and microscopy in inoculated tissue cultured sugarcane cultivars. Plant Science, v. 167, n. 5, p. 987–994.

SOLIAI M.M., *et al*. (2014). De novo genome assembly of the fungal plant pathogen Pyrenophora semeniperda. PLoS One, v. 9, n. 1, p. e87045.

SPERSCHNEIDER J., GARDINER D.M., DODDS P.N., TINI F., COVARELLI L., SINGH K.B., *et al*. (2016). EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. 210, 743–761.

SPERSCHNEIDER J., CATANZARITI A.M., DEBOER K., PETRE B., GARDINER D.M., SINGH K.B., *et al*. (2017). LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. Scientific reports,7, 44598.

STANKE M., MORGENSTERN B. (2004). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research, v. 33, p. W465–W467.

STOLL M., PIEPENBRING M., BEGEROW D., OBERWINKLER F. (2003). Molecular phylogeny of *Ustilago* and *Sporisorium* species (Basidiomycota, Ustilaginales) based on internal transcribed spacer (ITS) sequences 1.

STOLL M., BEGEROW D., OBERWINKLER F. (2005). Molecular phylogeny of *Ustilago*, *Sporisorium*, and related taxa based on combined analyses of rDNA sequences. Mycological Research, v. 109, n. 3, p. 342–356.

STUKENBROCK E.H., (2016). The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology, 106(2), p.104-112.

SUNDAR A.R., BARNABAS E.L., MALATHI P., VISWANATHAN R. (2012). A mini-review on smut disease of sugarcane caused by *Sporisorium scitamineum*. InTech Botany, p. 107–128.

TAMURA K., STECHER G., PETERSON D., FILIPSKI A., KUMAR S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30:2725-2729.

TANAKA S., SCHWEIZER G., RÖSSEL N., FUKADA F., THINES M., KAHMANN R. (2019). Neofunctionalization of the secreted Tin2 effector in the fungal pathogen Ustilago maydis. Nature microbiology, 4(2), p.251.

TANIGUTI L.M., SCHAKER P.D.C., BENEVENUTO J., PETERS L.P., CARVALHO G., PALHARES A., *et al*. (2015). Complete Genome Sequence of *Sporisorium scitamineum* and Biotrophic Interaction Transcriptome with Sugarcane. PLOS ONE, v. 10, n. 6, p. e0129318.

THIRUGNANASAMBANDAM P.P., HOANG N.V., HENRY R.J. (2018). The challenge of analyzing the sugarcane genome. *Frontiers in Plant Science*, *9*, p 616.

TUITE, J. (1969). Plant pathological methods. Fungi and bacteria. Plant pathological methods. Fungi and bacteria.

VÁNKY K. (2005). Taxonomic studies on Ustilaginomycetes - 25. Mycotaxon, v. 91, p. 217–272.

VOLLMEISTER E., SCHIPPER K., BAUMANN S., HAAG C., POHLMANN T., STOCK J., *et al*. (2012). Fungal development of the plant pathogen *Ustilago maydis*. FEMS Microbiology Reviews, v. 36, n. 1, p. 59–77.

WALKER B.J., ABEEL T., SHEA T., PRIEST M., ABOUELLIEL A., SAKTHIKUMAR S., *et al*., (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one, 9(11), e112963.

WATERHOUSE R.M., SEPPEY M., SIMÃO F.A., MANNI M., IOANNIDIS P., KLIOUTCHNIKOV G., *et al*., (2017). *BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol*, published online Dec 6, 2017.

YE Z., PAN Y., ZHANG Y., CUI H., JIN G., MCHARDY A.C., et al. (2017). Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome. DNA Res. 24, p. 635–648.

YIN Y., MAO X., YANG J., CHEN X., MAO F., XU Y., (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic acids research, 40(W1), W445-W451.

ZAMBANINI T., BUESCHER J.M., MEURER G., WIERCKX N., BLANK L.M., (2016). Draft genome sequence of *Ustilago trichophora* RK089, a promising malic acid producer. Genome Announc. 4:e00749-16.

ZUO W., ÖKMEN B., DEPOTTER J.R., EBERT M.K., REDKAR A., MISAS VILLAMIL J., *et al*. (2019).Molecular Interactions Between Smut Fungi and Their Host Plants. Annual review of phytopathology, 57, p.411-430.

SUPPLEMENTARY

TABLE 1 | Smut species and their ITS GenBank accession numbers.

Smut specie	GenBank Access number
Sporisorium panici-leucophaei	This work
Sporisorium scitamineum	CP010935
Sporisorium reilianum	AF045870
Ustilago hordei (O)	NSDP00000000
Ustilago hordei (B)	AY345003
Ustilago tritici syn. Tilletia tritici	NSHH00000000
Ustilago maydis	AY854090
Ustilago esculenta	AY345002
Ustilago trichophora	AY345009
Melanopsichium pennsylvanicum	AY740040

TABLE 2 | BUSCO *v3* output for *S. panici-leucophaei* v3 assembly.

BUSCO groups searched			
Complete BUSCOs		Missing and Fragmented BUSCOs	
1294 (97%)		41 (3%)	
Single copy	Duplicated	Missing	Fragmented
1293	1	26	15
Total BUSCO groups			
1335			

TABLE 3 | Annotation of *S. panici-leucophaei* mtDNA.

mtDNA annotation	Intron number
cob	4
tRNA-Ile(GAU)	-
cox3	-
tRNA-Ser (UGA)	-
tRNA-Tyr (GUA)	-
tRNA-Phe (GAA)	-
tRNA-Ala (UGC)	-
nad1	-
cox2	-
tRNA-Val (UAC)	-
tRNA-Arg (UCU)	-
atp6	-
atp8	-
cox1	7
tRNA-Ser (GCU)	-
tRNA-Lys (UUU)	-
tRNA-Gln (UUG)	-
tRNA-Thr (UGU)	-
tRNA-Glu (UUC)	-
nad4	-
tRNA-Met (CAU)	-
nad3 (copy 1)	-
nad (copy 2)	-
nad2	-
tRNA-Gly (UCC)	-
rps3	-
tRNA-Asn (GUU)	-
tRNA-Pro (UGG)	-
nad6	-
tRNA-Cys (GCA)	-
tRNA-Trp (CCA)	-
tRNA-Leu (UAG)	-
tRNA-Arg (UCG)	-
tRNA-Met (CAU)	-
rnpB	-
tRNA-His (GUG)	-
nad5	2
nad4L	-
tRNA-Asp (GUC)	-

Figure 1. Inoculated *D. insularis* (120 days after inoculation) emitting floral structure.