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

Molecular variability among Brazilian strains of the sugarcane smut pathogen and the genetic basis of host specialization in smut fungi

Juliana Benevenuto

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2017

Juliana Benevenuto
Biologist

Molecular variability among Brazilian strains of the sugarcane smut pathogen and the genetic basis of host specialization in smut fungi

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

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2017

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

Benevenuto, Juliana
Molecular variability among Brazilian strains of the sugarcane smut pathogen and the genetic basis of host specialization in smut fungi / Juliana Benevenuto. - - Piracicaba, 2017.
90 p.

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

1. Doenças do Carvão 2. Cana-de-açúcar 3. Efetores 4. Genômica 5. Adaptação ao hospedeiro I. Título

With all my love and gratitude

To my parents José Geraldo and Marta

To my brothers Ivo and Ivan

To my future husband Luis Felipe

ACKNOWLEDGEMENTS

This thesis was accomplished with the direct and indirect support of many people and institutions:

I acknowledge and thank the University of São Paulo (USP), specially the "Luiz de Queiroz College of Agriculture" campus (ESALQ). The opportunity to carry out my PhD studies at USP is the realization of a dream. From Minas Gerais, I have always admired the excellence of this institution.

I thank the Department of Genetics and the Graduate Program of Genetics and Plant Breeding, extended to professors, staff, and collegues, for the intellectual and practical support and good relationship.

I thank the São Paulo Research Foundation (FAPESP) for regular and BEPE schorlaships (grant numbers: 2014/21802-0 and 2016/03768-5). I also thank the National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES) for previous scholarships.

I sincerely thank my adviser, Prof. Claudia Barros Monteiro-Vitorello, for the support and outstanding guidance. Her combination of commitment and freedom of thought makes the lab a friendly and motivated environment to work. I am very grateful and honored to have been part of her team. "Thank you, Claudia, for your patience, all knowledge transmitted, friendly advices, and affection". I extend my gratitude to all members of the Genomics Lab for the kind and helpful environment since I joined the group: Elaine Vidotto, Patricia Schaker, Leila Peters, Natalia de Moraes, Lucas Taniguti, Giselle Carvalho, Mariana Marrafon, Suzane Saito, Daniel Longatto, Leandro de Souza, Pedro Beretta, Gustavo Crestana, Tatiane Shyton, Natalia Teixeira, Sintia Almeida, Andressa Bini, Natalia Araujo, Jackeline Borba, Marcela, Yago, Larissa. I especially thank Leila for her friendship and positivity, always with a word of encouragement; Paty and her husband Gabriel for the friendship and all good moments and trips togheter; Natalia Teixeira for all nice conversations; Lucas for being helpul and inspire me to learn bioinformatics; and Daniel for teaching me how to handle the fungus.

I thank Prof. Daniel Croll, Prof. Bruce A. McDonald, Sandra G. Weinreich and all members of the Plant Pathology Group, at ETH Zurich (Zurich, Switzerland), for receiving me and making me feel very welcome in the group. I specially thank Prof. Daniel Croll for the guidance, support with bioinformatics analyses, and all nice ideias and discussions that enriched this work. I also specially thank my dear colleagues with whom I shared the office, Norfarhan Mohd-Assaad and Fanny Hartmann. I take the opportunity to thank other very special people I met in Switzerland: Bethania Souza, Vladimir Souza, Marcela Terra, and Carolina Sardinha.

I thank Prof. Marie-Anne Van Sluys and Tatiana Corrêa for assistance in Sanger sequencing; PhD. Natalia Mielnichuk for the Argentine isolates; Prof. José Antonio Martinelli for field collection of smut pathogens; PhD. Eiko Kuramae for helping in the purchase and growing of smut pathogens from CBS bank; and the board of examiners for accepting the invitation.

I thank my lovely fiancé Luís Felipe Ventorim Ferrão for being by my side during the last five years, celebrating my achievements and giving me support in difficult moments. "Thank you for your love, confidence, and friendship. I love you!". I extend my gratitude to the Ventorim-Ferrão family for all support and affection and to Prof. Antonio Augusto Franco Garcia's lab, including here Amanda Avelar, for always receiving me so well.

My heartfelt thanks to my parents, José Geraldo Ferraz Benevenuto and Luzia Marta da Silva Benevenuto, and to my brothers, Ivo Benevenuto and Ivan Benevenuto, who watched me from distance. "Thank you for your endless love, encouragement, prayers, and lifelong efforts for me get here. *Eu amo vocês!*". I also express my gratitude to my entire family on behalf of my grandmothers Lili (*in memorian)* and Alzira. I also thank my friends from Visconde do Rio Branco and Viçosa for the love and friendship that remains despite the distance, specially, Fabíola (*in memorian*), Pollyana, Verônica, Fernanda, Polyanna, Andrea, Marcela, Giselle, Bárbara, David, and Iara.

I thank God for blessed my life with all such special and lovely people who have been made my walk lighter and happier.

SUMMARY

RESUMO

Variabilidade molecular entre isolados brasileiros do agente causal do carvão da cana-de-açúcar e a base genética da especialização ao hospedeiro

Fitopatógenos apresentam a habilidade de rapidamente suplantar os mecanismos de defesas da planta e adaptar-se a um novo hospedeiro. A (re)emergência de patógenos é uma das maiores preocupações na agricultura e na conservação de populações naturais. A rápida adaptação ao hospedeiro e a novos ambientes depende da variabilidade genética nas populações de patógenos. Apesar da importância da cana-de-açúcar para o agronegócio brasileiro e da persistência do patógeno *Sporisorium scitamineum*, o agente causal do carvão da cana-de-açúcar, na maioria das áreas canavieiras, estudos de variabilidade genética ainda não foram realizados para isolados brasileiros. Nos capítulos 1 e 2, estudos de variabilidade molecular foram realizados para isolados brasileiros e argentinos de *S. scitamineum*, usando marcadores moleculares (AFLP e telRFLP) e dados de sequenciamento (ITS e um gene candidato a efetor). Nenhum polimorfismo foi encontrado usando sequências ITS. Contrariamente, o marcador telRFLP gerou quase um *fingerprint* para cada linhagem. Dois grupos geneticamente distintos foram formados pela análise conjunta dos marcadores telRFLP e AFLP. Os dois grupos também foram formados pelos haplótipos obtidos pelo sequenciamento de um candidato a efetor. A presença de polimorfismos causando mutações não-sinônimas em um candidato a efetor pode acarretar em performances distintas em diferentes genótipos de cana-de-açúcar. *S. scitamineum* pertence à classe Ustilaginomycetes, a qual também abrange vários outros agentes causais de doenças do carvão. Apesar de filogeneticamente próximos e com estilo de vida similar, espécies de carvão apresentam uma faixa distinta e estreita de hospedeiros. Portanto, outro objetivo desta tese foi identificar a base genética da especialização ao hospedeiro por fungos causadores de carvão usando análises de genômica comparativa. No capítulo 3, os *loci* envolvidos na determinação do tipo de reação sexual (*mating-type*) foram caracterizados no genoma de *S. scitamineum* e comparados com sequências de outras espécies de carvão. Tranposons foram identificados como provável mecanismo de rearranjo cromossômico entre os *loci* de *mating-type*. Polimorfismos trans-específicos nos genes codificadores de feromônios e receptores sugerem o potencial de hibridização entre espécies de carvão. No capítulo 4, análises de genômica comparativa abrangendo nove espécies de carvão infectando hospedeiros distintos foram realizadas. A base genética da especialização ao hospedeiro em fungos causadores de carvão é complexa e parece envolver processos evolutivos de ganho/perda de genes e seleção positiva. Efetores espécie-específicos e sob seleção positiva são destacados como bons candidatos para serem caracterizados quanto ao papel que estabelecem na adaptação ao hospedeiro.

Palavras-chave: Doença do Carvão; Variabilidade; Efetores; *Mating-type*; Genoma; Adaptação ao hospedeiro; Genes orfãos; Seleção positiva

ABSTRACT

Molecular variability among Brazilian strains of the sugarcane smut pathogen and the genetic basis of host specialization in smut fungi

Plant pathogens have the ability to quickly overcome host resistance and shift to novel hosts. The (re)emergence of plant pathogens is a major concern in agriculture and in conservation of natural landscapes. The rapid adaptation to hosts and new environments depends on the genetic variability in pathogen populations. Despite of the importance of sugarcane for Brazilian agribusiness and the persistence of the smut pathogen *Sporisorium scitamineum* in most cropping areas, genetic variation studies are still missing for Brazilian isolates. In the chapters 1 and 2, molecular variability studies were performed for Brazilian and Argentine isolates of *S. scitamineum*, using molecular markers (AFLP, telRFLP) and sequencing (ITS and a candidate effector gene) strategies. No variation was found in ITS sequences. On the contrary, telRFLP marker generates almost a unique fingerprint for each strain. Two genetically distinct groups were formed by the joint analysis of the AFLP and telRFLP markers. The two groups were the same formed by haplotypes of a candidate effector gene. The presence of polymorphisms that causes non-synonymous mutations in a candidate effector gene potentially involved in the specific interaction with sugarcane may cause distinct performances on host genotypes. *S. scitamineum* is part of the highly diverse clade of Ustilaginomycetes fungi that includes several smut disease agents. Despite being phylogenetically close and present similar lifestyles, species of smut fungi have distinct and narrow host ranges. Hence, another objective in this thesis was to identify the genetic basis of host specialization in smut fungi using comparative genomics analyses. In chapter 3, the mating-type loci were described in *S. scitamineum* genome and compared among smut fungi. Transposable elements are the likely mechanism causing chromosomal rearrangements between mating-type loci. The presence of trans-specific polymorphisms at the genes encoding pheromone/receptor proteins suggests a hybridization potential among smut species. In the chapter 4, a broad comparative genomics analysis was performed among nine species of smut fungi infecting distinct hosts. The genetic basis of host specialization in smut fungi is complex and seems to involve a range of evolutionary processes, including gene gain/loss and episodic selection events. Species-specific effectors and positively selected genes will be good candidates for further characterization in regards to their role in host adaptation.

Keywords: Smut disease; Variability; Effectors; Mating-type; Genome; Host adaptation; Orphan genes; Positive selection

PREFACE

This thesis is presented as a requirement to obtain the degree of Doctor in Science at the "Genetics and Plant Breeding" Graduate Program, University of São Paulo (USP), campus "Luiz de Queiroz" College of Agriculture (ESALQ), Piracicaba, São Paulo, Brazil. The research described herein was conducted under the supervision of Professor Claudia Barros Monteiro-Vitorello between March 2013 and April 2017. The thesis was divided into four chapters that, in an overall view, aim to assess the molecular variability among Brazilian strains of *S. scitamineum* (the causal agent of sugarcane smut) and to understand the genetic basis of host specialization in smut fungi. The rapid evolution of plant pathogens seen at both inter- and intra-specific levels is what fascinates me and unifies the four chapters.

The first chapter entitled "Molecular variability and genetic relationship among Brazilian strains of the sugarcane smut fungus" was published in "FEMS Microbiology Letters" Journal. Part of the data was previously obtained by former students of the Genomics Group: Gislâine Vicente dos Reis, Daniel Prezotto Longatto, and Suzane Saito.

The second chapter is entitled "Polymorphic variant of a candidate effector potentially involved in the specific interaction between *Sporisorium scitamineum* and sugarcane". This data will be organized in a manuscript together with ongoing experiments of molecular characterization, expression time course and aggressiveness test of the two variants detected herein.

The third chapter entitled "Characterization of mating-type loci in *Sporisorium scitamineum* genome and implications in smut fungi evolution" was published as part of the paper "Complete Genome Sequence of *Sporisorium scitamineum* and Biotrophic Interaction Transcriptome with Sugarcane" in PLoS ONE Journal.

The fourth chapter entitled "The genetic basis of host specialization in smut fungi" was partially developed during my six months "Research Internship Abroad" (BEPE-FAPESP) under the supervision of Daniel Croll at ETH-Zurich, Zurich, Switzerland. We intend to submit it to "Genome Biology and Evolution" Journal.

GENERAL INTRODUCTION

1) Rapid evolution of fungal plant pathogens

Fungal plant pathogens have the ability to quickly overcome host resistance and shift to novel hosts. The (re)emergence of plant pathogens is a major concern in agricultural and natural landscapes (Fisher *et al*., 2012; McDonald and Stukenbrock, 2016). Pathogenic fungi usually have short generation time and produce billions of spores during an infection (Giraud *et al*., 2010). Such features allow the rapid creation of genetic variation by mutations (Giraud *et al*., 2010). High mutation rates increase the probability that a mutation affects virulence and pathogenicity genes, leading to higher fitness on resistant genotypes or novel hosts (McDonald and Linde, 2002).

Genomic features of fungal plant pathogens also appears as drivers of their rapid evolution (Croll and McDonald, 2012; Rafaelle and Kamoun, 2012; Dong *et al*., 2015). Plant pathogens genomes harbor large repertoires of effector genes potentially associated with virulence and pathogenecity. In multiple species, effectors are frequently found in the proximity of transposable elements (Raffaele and Kamoun, 2012; Dong *et al*., 2015). Such non-random association between effector genes and repetitive elements probably serve as cradle for adaptive evolution in fungal plant pathogens, allowing uneven patterns of evolution across the genome (Dong *et al*., 2015). Active transposable elements can accelerate the evolution by causing duplications, losses and disruptions of genes, generating new proteins by exon shuffling, regulating gene expression, or facilitating horizontal gene transfer (Wöstemeyer and Kreibich, 2002; Castanera *et al*., 2016).

Studying the genetic diversity at intra- and inter-specific levels can reveal the adaptive potential of pathogens to overcome host resistance and also provide insights into evolutionary mechanisms leading to specialization into novel hosts.

2) Co-evolution and host specialization

The evolutionary trajectory of pathogens is intimately associated with their hosts due to strong reciprocal selection pressures (Woolhouse *et al*., 2002). Pathogens have a negative effect on host fitness, triggering selection for enhanced defense mechanisms. Likewise, host defenses impose selection on pathogens for overcoming host immune responses. This antagonism governs coevolution, in which gene frequencies in one species determine the fitness of genotypes of the other species (Brown & Tellier, 2011).

The genetic basis of host-pathogen interactions was first proposed to be governed by gene-for-gene (GFG) relationships (Flor, 1956). The GFG hypothesis envisages that for each host resistance gene (R) there is a corresponding pathogen avirulence gene (Avr). In a classical receptor-ligand model, the protein products of R and Avr genes interact with each other triggering resistance response via hypersensitive cell death pathway. Currently, others mechanisms have also been associated with virulence, including fungal toxins, modulation of the host defense response, and transcriptional regulation (Poland *et al*., 2009). The precision of GFG interaction is considered the ultimate extreme of an evolutionary arms race (Thrall, 2001). To overcome the complex network of plant defense mechanisms and complete its life cycle, a successful pathogen probably accumulated a continuum of genetic changes throughout evolution (Kirzinger & Stavrinides, 2012).

The first line of plant defenses (basal immune system) involves the recognition of broadly conserved molecules in a wide range of microorganisms, named PAMPs (pathogen-associated molecular patterns). Well-known examples of fungal PAMPs are chitin and glucan residues that are detected by plant membrane-localized receptors, known as pattern recognition receptors (PRRs). The perception of cell wall damage by plants also activates basal defense signaling pathways (Bellincampi *et al*., 2014). The recognition of pathogen residues by PRRs induce "PAMP triggered immunity" (PTI) through the secretion of antifungal compounds, production of reactive oxygen species (ROS), callose deposition, protease inhibitors, inhibitors of plant cell wall degrading enzymes, chitinases and glucanases. PRRs are considered conserved and heritable defense mechanisms, allowing early detection of all potential pathogens. However, basal defenses are only partially effective at restricting pathogens (Bent and Mackey, 2007). Adapted pathogens are able to counteract PTI by secreting effector molecules into plant cells. Effectors can act either as an escape of host recognition, suppressor of host immune response, manipulator of host cell physiology, or a toxin that kill the host cell. However, effectors can activate a second layer of plant immune response, termed "effector triggered immunity" (ETI). Effectors that activate ETI are known as an avirulence (Avr) factors and are recognized by specific plant resistance proteins (R proteins) as proposed in the GFG concept. PTI and ETI triggered similar defense responses, although in ETI the defense responses are enhanced in timing and amplitude. The ETI leads to hypersensitive cell death response (HR) that arrests the pathogen growth at the infection site, resulting in an incompatible interaction or resistance. The HR is effective against biotrophic pathogens by restricting pathogen access to nutrients and activating salicylic acid-dependent signaling. Necrotrophic pathogens promote HR-like cell death and jasmonic acid and

ethylene-dependent pathways are major regulators of defense responses against necrotrophics (Glazebrook, 2005). In contrast to the conserved molecules involved in PTI, effectors and R proteins are highly variable, waging an arms race (Jones and Dangl, 2006). Fast evolutionary changes in effectors (Avr) genes make them unrecognizable by the host R genes, resulting in a compatible interaction, or disease (Sonah *et al*., 2016). Conventional mutations, pseudogenization, expansion/contraction of gene families, gene gain/loss, and also genetic and epigenetic changes in controlling the effector gene expression state can promote gain of virulence (Gijzen *et al*., 2014).

The current cost-effectiveness of sequencing technologies has opened the way for the whole-genome sequencing of many organisms. Comparative genomics tools have been applied widely across the fungal pathogens, providing insights into mechanisms of pathogenicity, lifestyle, and genome plasticity (Plissonneau *et al*., 2017). A particular focus has been the repertoire of genes encoding effectors. Effectors genes are predicted as encoding small, cysteine-rich and secreted proteins. Besides their role in pathogenicity and virulence, effectors are also associated with the host range of pathogens (Djamei *et al*., 2011; Feldbrügge *et al*., 2013).

Due to coevolution, pathogens are expected to evolve to higher degrees of host specialization (Antonovics *et al*., 2012) and, therefore, phylogenetically constrained host ranges (Gilbert and Webb, 2007). As a consequence of the intimate interaction established with their hosts, in general, biotrophic pathogens have a narrow host range (Oliver and Ipcho, 2004). Specialist pathogens evolved host-specific adaptations that enable them to infect, defeat of host defenses, uptake nutrients, multiply and reproduce within the host (Haueisen and Stukenbrock, 2016). Hence, specialization onto hosts probably involve different genetic and evolutionary basis in each fungal pathogen.

Smut fungi are suitable pathogens to investigate host specialization mechanisms. Most of smut species infect only a single or a small number of susceptible hosts (Begerow *et al*., 2004). The host specificity seen in the field are so striking that, for over 200 years, host associations were used as an "Ecological Species Criterion" for smut classification (Cai *et al*., 2011).

3) Smut diseases

True smut fungi belong to Ustilaginomycetes class into the Basidiomycota phylum and comprise more than 1,650 smut species (Toh and Perlin, 2016; Cai *et al*., 2011). Smut

hosts are distributed over many angiosperm clades, although most occurs on monocots, especially on Poaceae family members (Begerow *et al*., 2004). Many agronomically important crops such as cereals, sugarcane, and forage grasses are affected (Martínez-Espinoza *et al*., 2002). Losses range from negligible to significant proportions (Christensen, 1963; Quijano *et al*., 2016; Sundar *et al*., 2012). In general, smut pathogens are found along the geographic distribution of their hosts (Begerow *et al*., 2014). The dispersion mainly occurs through the teliospores carried by wind, water or animal vectors.

Smuts are characterized by the biotrophic growth *in planta* culminating with the production of a sooty dark brown mass of teliospores (Bakkeren *et al*., 2008; Morrow and Fraser, 2009). Under proper environmental conditions, the diploid teliospores germinate and, thereafter, entail meiosis forming haploid sporidia, which grow as saprophytic budding-like yeasts (Figure 1). In the smut life cycle occurs a switch between the haploid saprophytic yeast form to the dikaryotic infective hyphae in a process triggered by mating reaction between two compatible sporidia. This morphogenetic transition and sexual reproduction are required to smut pathogenesis, since outside the host, sporogenesis does not occur (Bakkeren *et al*., 2008). After penetrating the plant surface, most of smut species systemically colonize the host plant and intense proliferation occurs close to meristematic tissue (Piepenbring, 2009). In general, the plants remain asymptomatic until the fungal sporogenesis (Martínez-Espinoza *et al*., 2002; Brefort *et al*., 2009; Morrow and Fraser, 2009). Sporogenesis mainly ocurrs in ovaries, flowers, or in entire inflorescences where there is massive fungal proliferation, followed by karyogamy, hyphae fragmentation, pigment deposition and formation of billions of diploid teliospores that reinitiate the cycle (Figure 1). In addition to the host range, smut species also differ in secondary symptoms, local of infection and local of sporogenesis (Figure 2) (Table 1).

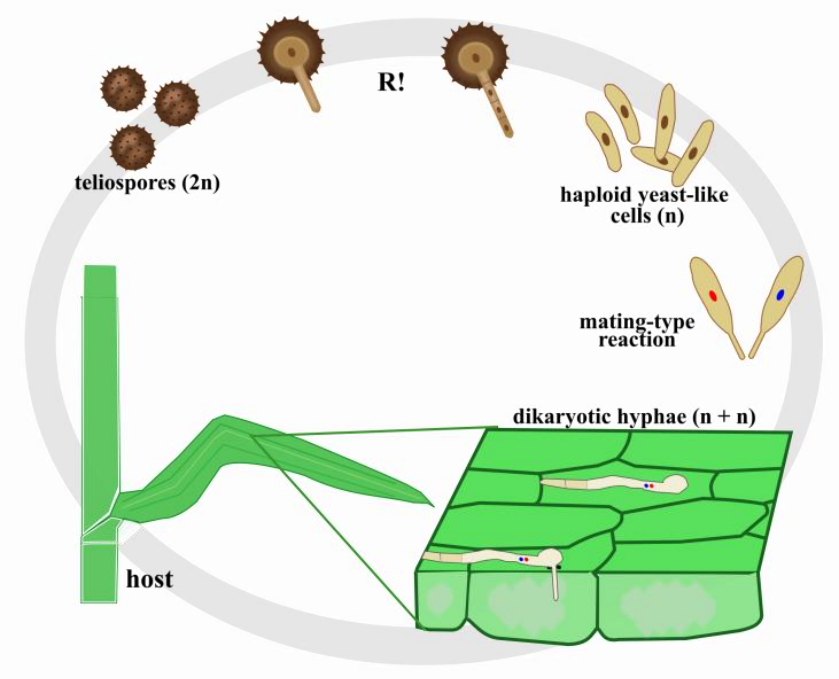

Figure 1. General representation of smut fungi lifecycle. The three phases are shown: diploid teliospores, haploid yeast-like cells, and dikaryotic infective hyphae. Meiosis is shown as R!. Opposite mating-type cells are indicated by red and blue nuclei.

Table 1. Differences among smut species

Smut species	Host range	Penetration	Colonization	Sporogenesis	Secondary symptoms	References
U. maydis	maize, teosinte (common smut)	all aerial parts	local	infected tissue	Tumor, anthocyan accumulation	Brefort *et al*., 2009; Bölker, 2001.
U. hordei	barley, oat (covered smut)	coleoptile	systemic	ovary	stunting	Hu *et al*., 2002; Martínez-Espinosa *et al*., 2002.
U. tritici	wheat, barley (loose smut)	coleoptile	systemic	ovary	chlorotic flecks on leaf	Jones, 1999;Batts, 1955.
U. esculenta	*Zizania latifolia* (Manchurian wild rice)	young seedlings/ propagated with plant rhizome	non-systemic	3-4 apical internodes	tumor , prevents the host plant from producing inflorescences	Chung and Tzeng, 2004; Zhang *et al*., 2012.
U. trichophora	*Echinochloa* genus	seedlings	na*	inflorescences and vegetative parts	tumor and malformation at inflorescences and vegetative parts	Zhang *et al*., 2013.
S. reilianum	maize, sorghum (head smut)	roots	systemic	ears and tassel	chlorotic flecks on leaf, anthocyan accumulation on stem, stunting, phyllody, greater number of ears,tillering in sorghum	Zuther *et al*., 2012; Martinez *et al*., 2002.
S. scitamineum	sugarcane	young buds	systemic	apical meristem	tillering, narrow leaves and culms, increased fiber and reduced sucrose content	Sundar *et al*., 2012; Taniguti *et al*., 2015.
S. sorghi	sorghum (covered kernel smut)	coleoptile	systemic	ovary	na*	Moharam *et al*., 2012; Narayanasamy, 2006.
M. pennsylvanicum	*Persicaria spp*. (dicot)	na*	na*	inflorescences and vegetative parts	tumor in terminal floral racemes and vegetative tissue	Halisky and Barbe, 1962; Sharma *et al*., 2014.

*na=not available

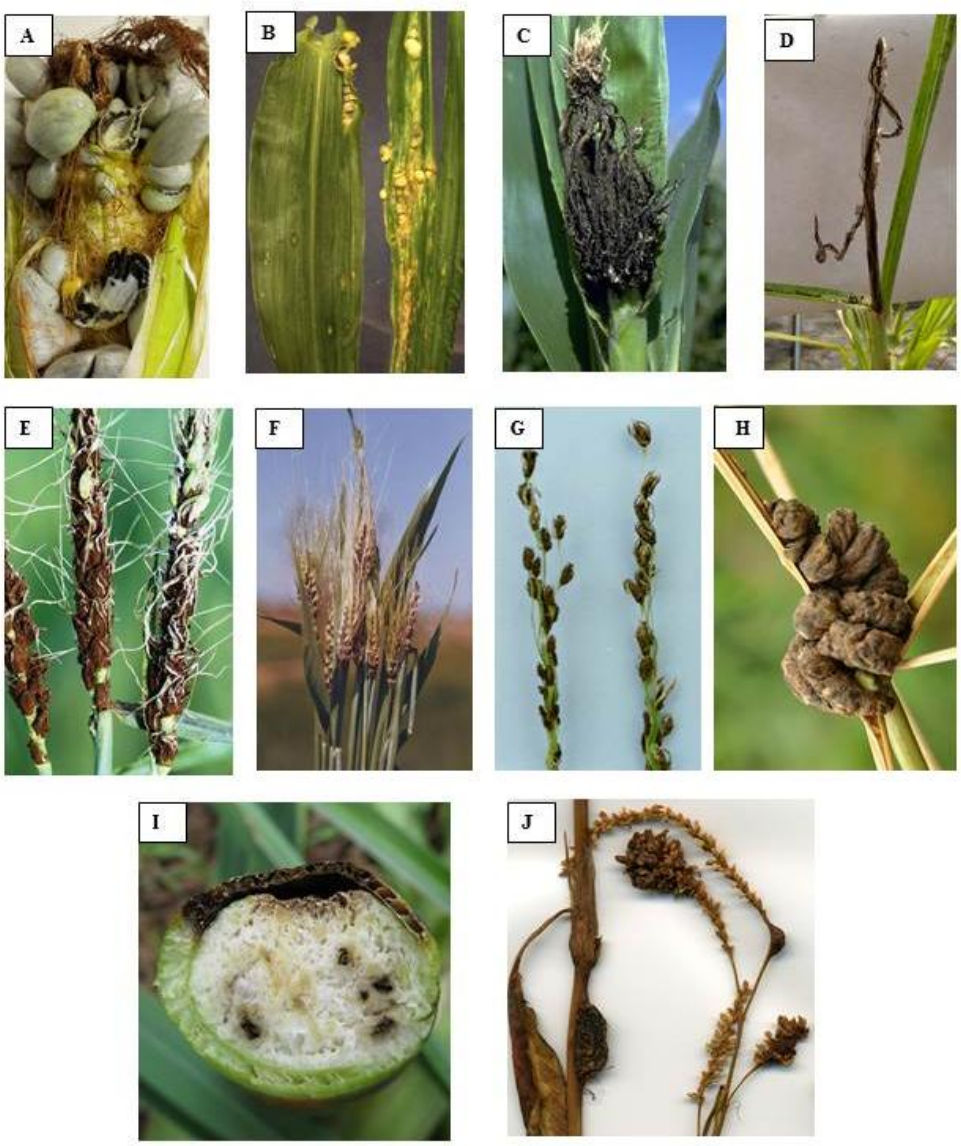

Figure 2. Smut diseases caused by different species. **A)** Common corn smut caused by *U. maydis*. **B)** Symptoms of leaf infection by *U. maydis*. Image by Margaret McGrath, Cornell University, Bugwood.org **C)** Corn head smut caused by *S. reilianum*. Image from R.L. Croissant, Bugwood.org. **D)** Sugarcane smut caused by *S. scitamineum*. **E)** Loose smut of wheat caused by *U. tritici*. Image from Donald Groth, Louisiana State University AgCenter, Bugwood.org. **F)** Coved smut of barley caused by *U. hordei*. Image from Rikhab Raj-Bhansali, CAZRI, Bugwood.org. **G)** Coved smut of oats caused by *U. hordei*. Image from Clemson University - USDA Cooperative Extension Slide Series, Bugwood.org. **H)** *U. trichophora*, the causal agent of smut in *Echinochloa* genus. Image from OneDrive, added by Mirosław Wantoch-Rekowski. **I)** *U. esculenta* causing smut on *Zizania latifolia*. Image from Shivas *et al*. (2013), at Smut Fungi of Thailand website: http://collections.daff.qld.gov.au/thaismutfungi, accessed on 23/02/2017. **J)** *Melanopsichium pennsylvanicum* on *Persicaria lapathifolia*. Image from Shivas *et al*. http://collections.daff.qld.gov.au/web/key/smutfungi/Media/Html/melanopsichiumpennsylvanicum.html, accessed on 23/02/2017. No additional permissions were indicated in the websites for using the images for non-commercial purposes.

4) Sugarcane smut disease

Sugarcane (*Saccharum spp*.) is an important agricultural crop mostly due its high sucrose accumulation capacity. Sugarcane has been responsible for 70% of the world's sugar production and is an increasingly source of alcohol-based fuel (McCormick *et al*., 2009). Sugarcane belongs to Poaceae family and modern varieties are interspecific hybrids from natural and artificial crosses between species of the genus *Saccharum*, mainly involving crosses between *S. officinarum* and *S. spontaneum* (Souza *et al*., 2011). Sugarcane has a complex aneu-polyploid genome (2n = 100-130) (D'Hont, 2005; Souza *et al*., 2011), hindering the understanding of the genetic architectures of agronomically important traits (Cheavegatti-Gianotto *et al*., 2011). Due to its genome complexity, sugarcane commercial varieties are vegetatively propagated.

Sugarcane is widely cultivated in tropical and subtropical regions of the world. In 2014, the world production of sugarcane was about 1.8 billion tons on a harvest area of 27 million hectares (FAOSTAT, 2014). Brazil is the largest sugarcane producer, being responsible for around 40% of the world's production (FAOSTAT, 2014). The State of São Paulo is the largest national producer, responsible for 55% of Brazil's production (UNICA, 2016).

Sporisorium scitamineum (Syd.) [Piepenbring *et al*. (2002) (Syn: *Ustilago scitaminea* H. and P. Sydow)] is the causal agent of sugarcane smut and one of the major threats to the culture. The disease is present in almost all sugarcane growing area (Comstock, 2000; Croft & Braithwaite, 2006). The main symptom of the disease is the emergence of a long black whip-like structure (*sori* containing teliospores) in the shoot apical meristem of the plant. When the infection occurs in young plants, abundant tillering and narrow leaves are reported, which makes the cane appears ''grass-like" (Sundar *et al*., 2012). The disease leads to increased fiber and reduced sucrose content, causing significant losses in cane tonnage and juice quality (Tokeshi and Rago, 2005; Wad *et al*., 2016).

Yield losses vary widely, depending on the environmental conditions, the pathogen races, and the tolerance level of the sugarcane varieties (Sundar *et al*., 2012).Sugarcane varieties resistant to smut disease appear to be sustainable, but there is no smut immunity in sugarcane genotypes (CanaOnline, 2014). In addition, the complexity of sugarcane genome makes it difficult to identify disease specific resistance genes (Rott *et al*., 2013). Few breakdowns have been reported for smut diseases and the presence of pathogen races is still controversial (Ferreira and Comstock,1989). Differences in sugarcane genotypes responses to

distinct fungal populations was found in many localities, but strong evidence for distinct races was observed only in Taiwan (Grisham, 2001).

A major concern in sugarcane breeding is the highly homogenous environment provided by sugarcane fields with clonal varieties, long-term monoculture and uniform agronomic practices. If a disease outbreak occurs due to new virulent isolate, these homogenous environments will be conducive for quickly pathogen dispersal, leading to the devastation of large areas. Another raising concern is the introgression of *S. spontaneum* germoplasm in breeding programs for energy cane. The wild species *S. spontaneum* is not resistant to smut and is considered a collateral host and a major reservoir of inoculum in India (Braithwaite *et al*., 2004; Jose *et al*., 2016; Srinivasan and Chenulu, 1953). It is also important to point out that climate changes and the prohibition of burn-before-harvesting management of the crop in Brazil can also have large effects on the disease epidemiology.

Monitoring the genetic variability in pathogen populations is very important for optimize breeding strategies and screening cultivar genotypes for resistance, and, thus, stay ahead of pathogens.

References

Antonovics J, Boots M, Ebert D, Koskella B, Poss M, Sadd BM (2012). The origins of specificity by means of natural selection: Evolved and nonhost resistance in host-pathogen interactions. *Evolution* **6**:1–9.

Bakkeren G, Kämper J, Schirawski J (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. *Fungal Genetics and Biology* **45**:S15-21.

Batts CCV (1955). Loose smut (*Ustilago tritici* (Pers.) Rostr) of wheat: physiologic specialization and reaction of varieties in England. *Annals of Applied Biology* **43**(4): 533-537.

Begerow D, Goker M, Lutz M,Stoll M (2004). On the evolution of smut fungi on their hosts. In Frontiers in Basidiomycote Mycology (R. Agerer, M. Piepenbring & P. A. Blanz, eds): 81–98. IHW-Verlag, Eching.

Begerow D, Schafer AM, Kellner R, Yurkov A, Kemler M., *et al*. (2014). Ustilaginomycotina. In: The Mycota, Systematics and Evolution (McLaughlin DJ, Spatafora JW, eds). 2nd edn Berlin: Springer. v. 7A: 295–329.

Bellincampi D,Cervone F, Lionetti V. Plant cell wall dynamics and wall-related susceptibility in plant–pathogen interactions. *Frontiers in Plant Science* **5** (228):1-8.

Bent AF, Mackey D (2007). Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Annual Review of Phytopathology **45**:399-436.

Bölker M. 2001. *Ustilago maydis*—a valuable model system for the study of fungal dimorphism and virulence. *Microbiology* **147**: 1395–1401.

Braithwaite KS, Bakkeren G, Croft BJ, Brumbley SM (2004). Genetic variation in a worldwide collection of the sugarcane smut fungus *Ustilago scitaminea*. *Proceedings of the Australian Society of Sugar Cane Technology* **26**: 1–9.

Brefort T, Doehlemann G, Mendoza-Mendoza A, Reissmann S, Djamei A, Kahmann R (2009). *Ustilago maydis* as a Pathogen. Annual Review of Phytopathology. 47:423-45.

Brown JK, Tellier A. Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology* **49**:345-367

Cai L, Giraud T, Zhang N, Begerow D, Cai D, Shivas RG (2011). The evolution of species concepts and species recognition criteria in plant pathogenic fungi. *Fungal Diversity* **50:** 12.

CanaOnline (2014). Não existe variedade imune ao carvão. Avaiable at: http://www.canaonline.com.br/conteudo/-nao-existe-variedade-imune-ao-carvao.html#.WK7zFm_yt0w. Accessed: 23/02/2017.

Castanera R, López-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.

Cheavegatti-Gianotto A, Abreu HMC, Arruda P, Bespalhok-Filho JC, Burnquist WL, Creste S, Ciero L, *et al*. (2011). Sugarcane (Saccharum X officinarum): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. *Tropical Plant Biology* **4**(1): 62–89.

Christensen JJ (1963). Corn smut caused by *Ustilago maydis*. Monograph 2, St. Paul (Minnesota): American Phytopathological Society :1–40.

Chung KR, Tzeng DD (2004). Biosynthesis of indole-3-acetic acid by the gall-inducing fungus *Ustilago esculenta*. *Journal of Biological Sciences* **4**:744–750.

Comstock, J.C. (2000) Smut. In A guide to sugarcane diseases Ed. P. Rott, J.C. Comstock, B.J. Croft and A.S. Saumtally. CIRAD/ISSCT, Montpellier.

Croll D, McDonald BA (2012) The Accessory Genome as a Cradle for Adaptive Evolution in Pathogens. *PLoS Pathogens* **8**(4): e1002608.

Croft, B.J. & Braithwaite, K.S. (2006). Management of an incursion of sugarcane smut in Australia. Australasian Plant Pathology 35(2) 113-122.

D'Hont A, 2005. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenetic and Genome Research* **109**: 27–33.

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):395-398.

Dong S, Raffaele S, Kamoun S. (2015). The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development* **35**:57-65.

FAOSTAT (2014). Food and agriculture data. Avaiable at: http://www.fao.org/faostat/ . Accessed: 23/02/2017.

Feldbrügge M, Kellner R, Schipper K (2013). The biotechnological use and potential of plant pathogenic smut fungi. *Applied Microbiology and Biotechnology* **97**(8):3253-3265.

Ferreira SA, Comstock JC, 1989. Smut. In: Ricaud C, Egan BT, Gillaspie AG, Hughes CG. (eds.), Diseases of Sugarcane. Elsevier, Amsterdam, pp. 211–229.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**: 7393.

Flor HH (1956) The Complementary Genic Systems in Flax and Flax Rust. *Advances in Genetics* **8**: 29-54.

Gijzen M, Ishmael C,Shrestha SD (2014). Epigenetic control of effectors in plant pathogens. *Frontiers in Plant Science* **5**: 638.

Gilbert GS, Webb CO (2007). Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences of the United States of America* **104**(12):4979-4983.

Giraud T, Gladieux P, Gavrilets S (2010). Linking emergence of fungal plant diseases and ecological speciation. *Trends in Ecology & Evolution* **25**(7): 387–395.

Glazebrook J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annual Review of Phytopathology **43**:205-227.

Grisham MP (2001). An international project on genetic variability within sugarcane smut. *Proceedings of the International Society of Sugarcane Technologists* **24**: 459-461.

Halisky PM, Barbe GD (1962). A Study of *Melanopsichium pennsylvanicum* Causing Gall Smut on *Polygonum*. *Bulletin of the Torrey Botanical Club* 89(3):181-186.

Haueisen J, Stukenbrock EH (2016). Life cycle specialization of filamentous pathogens - colonization and reproduction in plant tissues. *Current Opinion in Microbiology* **32**:31-37.

Hu, G.G., Linning, R., and Bakkeren, G. 2002. Sporidial mating and infection process of the smut fungus, *Ustilago hordei*, in susceptible barley. *Canadian Journal of Botany* **80**: 1103–1114.

Jones JDG, Dangl JL (2006). The plant immune system. *Nature* **444**: 323-329.

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.

Jose RC, Louis B, Goyari S, Waikhom SD, Handique PJ, Talukdar NC (2016). Biotrophic interaction of *Sporisorium scitamineum* on a new host - *Saccharum spontaneum*. *Micron* **81**: 8–15.

Kirzinger MW, Stavrinides J (2012). Host specificity determinants as a genetic continuum. *Trends in Microbiology* **20**(2):88-93.

Martinez C, Roux C, Jauneau A, Dargent R (2002). The biological cycle of *Sporisorium reilianum f.sp. zeae*: an overview using microscopy. Mycologia **94**: 505–514.

Martinez-Espinoza AD, Garcia-Pedrajas MD, Gold SE (2002). The Ustilaginales as plant pests and model systems. *Fungal Genetics and Biology* **35**(1):1-20.

McCormick AJ, Watt DA, Cramer MD (2009). Supply and demand: sink regulation of sugar accumulation in sugarcane. Journal of Experimental Botany **60**(2): 357-364.

McDonald BA, Linde C (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**:349-379.

McDonald BA, Stukenbrock EH (2016). Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philosophical Transactions of the Royal Society B* **371**: 20160026.

Moharam MH, Leclerque A, Koch E (2012). Cultural characteristics of *Sporisorium sorghi* and detection of the pathogen in plant tissue by microscopy and PCR. *Communications in agricultural and applied biological sciences* **77**(3):297-309.

Morrow CA, Fraser JA (2009). Sexual reproduction and dimorphism in the pathogenic basidiomycetes. *FEMS Yeast Research* **9**(2):161-177.

Narayanasamy P (2006). Disease development and symptom expression. In Postharvest Pathogens and Disease Management. John Wiley & Sons, New Jersey, USA. p 117-174.

Oliver RP, Ipcho SV. Arabidopsis pathology breathes new life into the necrotrophs-vs.-biotrophs classification of fungal pathogens. *Molecular Plant Pathology* **5**(4):347-352.

Piepenbring M (2009). Diversity, Ecology, and Systematics of Smut Fungi. In Tropical Biology and Conservation Management, v.6: Phytopathology and Entomology. Del Claro K, Oliveira PS, Rico-Grey V. eds., EOLSS Publishers/UNESCO, Oxford, United Kimgdom.

Piepenbring M, Stoll M, Oberwinkler F (2002). The generic position of *Ustilago maydis*, *Ustilago scitaminea*, and *Ustilago esculenta* (Ustilaginales). Mycological Progress 1: 71–80.

Plissonneau C, Benevenuto J, Mohd-Assaad N, Fouché S, Hartmann FE, Croll D (2017). Using Population and Comparative Genomics to Understand the Genetic Basis of Effector-Driven Fungal Pathogen Evolution. *Frontiers in Plant Science* **8**:119.

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009). Shades of gray: the world of quantitative disease resistance. *Trends in Plant Science* **14**(1): 21–29.

Quijano CD, Wichmann F, Schlaich T, Fammartino A, Huckauf J, Schmidt 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* **11**: 90-98.

Raffaele S, Kamoun S (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* **10**(6):417-430.

Rott PC, Girard J-C, Comstock JC (2013). Impact of pathogen genetics on breeding for resistance to sugarcane diseases. Proceedings International Society of Sugar Cane Technologists, 28: 1-11.

Sharma R, Mishra B, Runge F,Thines M. Gene Loss Rather Than Gene Gain Is Associated with a Host Jump from Monocots to Dicots in the Smut Fungus *Melanopsichium pennsylvanicum*. *Genome Biology and Evolution* **6**(8): 2034-2049.

Sonah H, Deshmukh RK, Bélanger RR (2016). Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges. *Frontiers in Plant Science* **7**: 126.

Souza GM, Berges H, Bocs S, Casu R, D'Hont A, Ferreira JE, Henry R, Ming R, Potier B, Van Sluys MA, Vincentz M, Paterson AH, 2011. The sugarcane genome challenge: Strategies for sequencing a highly complex genome. *Tropical Plant Biology* **4**: 145–156.

Srinivasan KV, Chenulu VV (1953). A preliminary study of the reaction of *Sacharum spontaneum* variants to red rot, smut, rust and mosaic. *Proceedings International Society of Sugarcane Technologists Congress* **9**: 1097–1107.

Sundar AR, Barnabas EL, Malathi P, Viswanathan R (2012). A mini-review on smut disease of sugarcane caused by *Sporisorium scitamineum*. In Botany, ed. Mworia J. Rijeka: InTech Publisher, 109–128.

Taniguti LM, Schaker PDC, Benevenuto J, Peters LP, Carvalho G, Palhares A, *et al*. (2015). Complete Genome Sequence of *Sporisorium scitamineum* and Biotrophic Interaction Transcriptome with Sugarcane. *PLoS ONE* **10**(6): e0129318.

Thrall PH, Watkin ELJ, Burdon JJ (2001). Coevolution: plant–Microorganism. *Encyclopedia of Life Sciences*, John Wiley & Sons.

Toh SS, Perlin MH (2016). Resurgence of Less-Studied Smut Fungi as Models of Phytopathogenesis in the Omics Age. Phytopathology **106**(11):1244-1254.

Tokeshi H, R.A., 2005. Doenças da cana-de-açúcar, in: Kimati H., Amorim L., Rezende J.A.M., Bergamin Filho A., C.L.E.A. (Ed.), Manual de Fitopatologia: Doenças Das Plantas Cultivadas. Agronômica Ceres, São Paulo, pp. 185–196.

UNICADATA (2016). União da Industria de Cana-de-açúcar. Avaiable at: http://www.unicadata.com.br . Accessed: 23/02/2017.

Wada AC, Anaso AB, Bassey MS (2016). Sugar Cane Whip Smut (*Sporisorium scitamineum* Syd) Caused Field Sucrose and Juice Quality Losses of Two Sugar Cane Varieties in Nigeria. *International Journal of Plant & Soil Science* **10**(4): 1-11.

Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32**(4):569-77.

Wöstemeyer J, Kreibich A. Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Current Genetics* **41**(4):189-198.

Zambanini T, Buescher JM, Meurer G, Wierckx N, Blanka LM (2016). Draft Genome Sequence of Ustilago trichophora RK089, a Promising Malic Acid Producer. *Genome Announcements* 4(4): e00749-16.

Zhang JZ, Chu FQ, Guo DP, Hyde KD, Xie GL (2012).Cytology and ultrastructure of interactions between *Ustilago esculenta* and *Zizania latifolia*. *Mycological Progress* **11**: 499–508.

Zhang JZ, Guan P, Tao G, Ojaghian MR, Hyde KD (2013). Ultrastructure and phylogeny of *Ustilago coicis* . *Journal of Zheijang University SCIENCE B**14***(4): 336–345

Zuther K, Kahnt J, Utermark J, Imkampe J, Uhse S, Schirawski J. 2012. Host specificity of *Sporisorium reilianum* is tightly linked to generation of the phytoalexin luteolinidin by *Sorghum bicolor*. *Molecular Plant-Microbe Interactions* **25**:1230–1237.

CHAPTER 1: Molecular variability and genetic relationship among Brazilian strains of the sugarcane smut fungus

ABSTRACT

Sporisorium scitamineum is the biotrophic fungus that causes sugarcane smut disease. Despite of the importance of sugarcane for Brazilian agribusiness and the persistence of the pathogen in most cropping areas, genetic variation studies are still missing for Brazilian isolates. In this study, sets of isolates were genotyped using two molecular markers (AFLP and telRFLP) and ITS sequencing. Twenty-two whips were collected from symptomatic plants in cultivated sugarcane fields of Brazil. A total of 41 haploid strains of compatible mating types (A and B) were selected from individual teliospores and used for molecular genetic analyses. telRFLP and ITS analyses were expanded to six Argentine isolates, a Brazil's neighbor country where the sugarcane smut disease was first recorded in America. The genetic distance matrix among strains was not significantly correlated with any of the variables tested (spatial distribution, collection date, and breeding program origin). Genetic relationship among strains suggests the human-mediated dispersal of *S. scitamineum* within the Brazilian territory and between the two neighboring countries. Two genetically distinct groups were defined by the combined analysis of AFLP and telRFLP. The opposite mating-type strains derived from a single teliospore were clustered together into these main groups, but had not always identical haplotypes. No polymorphism was detected by using ITS sequences. On the contrary, telRFLP marker generates almost a unique fingerprint for each strain. Moreover, telRFLP markers analyzed over two generations of selfing and controlled outcrossing confirmed the potential for emergence of new genetic variants and occurrence of recombination, which are relevant events for evolution of virulence and environmental adaptation.

Keywords: *S. scitamineum*; telRFLP; AFLP; ITS; Brazil Argentina

1) INTRODUCTION

Sporisorium scitamineum is the causal agent of sugarcane smut, a disease distributed worldwide (Comstock & Lentini, 2005). The fungus presents three genetically and morphologically distinct phases during its life cycle: haploid yeast-like sporidia, dikaryotic hyphae and diploid teliospores (Fig. 1). Infection culminates with the outgrowth of a whip-like structure from shoot primary meristem, characterizing the main symptom of the disease (Comstock, 2000). Within this structure, teliospores develop as a result of karyogamy and hyphae fragmentation. Under appropriate environmental conditions, the teliospores germinate, undergo meiosis, and produce haploid sporidia that grow saprophytically as yeast-like cells and *in vitro*. To infect the host, a combination of two haploid sporidia from opposite mating-types is needed to form the dikaryotic infective hyphae (Alexander & Srinivasan, 1966).

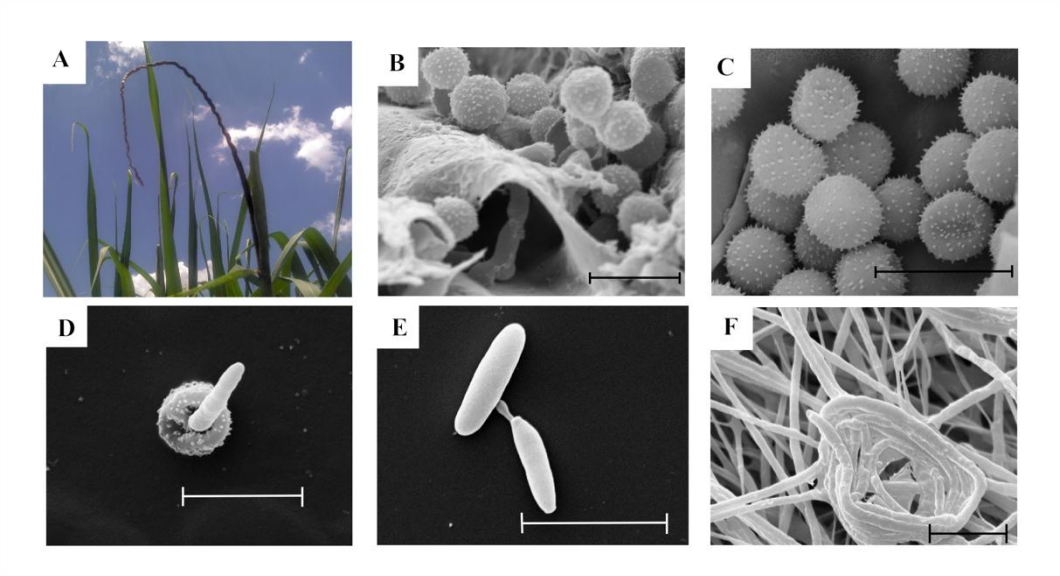

Fig.1. Life cycle stages of *S. scitamineum*. **(A)** Whip-like structure obtained from IAC982053 sugarcane cultivar; **(B)** and **(C)** Scanning electron micrographs of the whip covered by a silvery membrane sheltering teliospores; **(D), (E)** and **(F)** growth in culture medium of germinating teliospore, haploid yeast-like cells, and hyphae, respectively. Scale bar, 10 μM.

The complete genome sequence of *S. scitamineum* (strain SSC39A) confirmed the bipolar mating system by showing that the two sex-determining *loci* (*a* and *b*) are physically linked at a distance of 59 kbp apart on chromosome 2 (Taniguti *et al*., 2015). In general, the bipolar system favors a higher level of selfing compared with the tetrapolar system, leading to elevated homozygosity in the fungal population (Fraser & Heitman, 2003), reducing the reassortment of genetic variability, and thus the potential of the pathogen to adapt to evolutionary changes of the host (Kaltz & Shykoff, 1999). On the other hand, selfing could be selectively advantageous in cases of limited partner availability, ensuring reproductive success (Billiard *et al*., 2012). The advantages of inbreeding or outbreeding may depend on the ecological niche occupied by the fungus (Bakkeren *et al*., 2008) and their occurrence is also dependent on the number and frequency of alternate alleles at the mating-type *loci* (Giraud *et al*., 2008).

Many genetic studies have been conducted on global and local levels, using different strategies to assess *S. scitamineum* molecular variability and virulence degree (Bhuiyan *et al*., 2015; Braithwaite *et al*., 2004; Fattah *et al*., 2010; Luzaran *et al*., 2012; Que *et al*., 2012; Raboin *et al*., 2007; Rago *et al*., 2009; Singh *et al*., 2005; Xu *et al*., 2004, 2014; Zhang *et al*., 2015). In an overall view, low levels of genetic variation were found in American and African isolates while high levels were found in Southeast Asian populations. Southeast Asia was predicted as the center of origin of *S. scitamineum* from where a single lineage was dispersed

to other continents through exchange of infected plant material for breeding purposes (Braithwaite *et al*., 2004; Raboin *et al*., 2007). The complete homozygosity of 17 co-dominant markers for all fungal isolates analyzed by Raboin *et al*. (2007), indicates that selfing is the predominant reproductive mode and that the current population structure is markedly homogeneous.

Despite the importance of Brazil as the largest producer of sugarcane in the world and the continuous maintenance of the fungus in cultivated areas, a detailed survey of the genetic variability of this pathogen in Brazilian fields is still missing. In order to better understand the molecular variation and the genetic relationship among Brazilian strains of *S. scitamineum* for further improve breeding strategies for smut resistance, we used a combination of molecular approaches. Each technique accesses different genomic regions: AFLPs (Amplified Fragment Length Polymorphisms) are randomly distributed throughout the genome (Mueller and Wolfenbarger, 1999); telRFLPs (telomere-associated Restriction Fragment Length Polymorphisms) are located at chromosome termini (Wu *et al*., 2009); and ITS (Internal Transcribed Spacers) are located at ribosomal DNA and is a popular region for variability studies (Nilsson *et al*., 2008). In addition, we also investigate:1) the genetic relationship between Brazilian and Argentine isolates, since Argentina is a Brazil's neighbor country where the sugarcane smut disease was first recorded in America in 1940 (Ferreira & Comstock, 1989); 2) the presence of polymorphisms between two sexual compatible haploid strains derived from a single teliospore; 3) the usage of distinct restriction enzymes from those used by Braithwhite *et al*.(2007) in AFLP-based genotyping; 4) the stability of telRFLP profiles over generations.

2) MATERIAL AND METHODS

2.1) Ethics Statement

Smut whips were collected from 22 plants growing at sugarcane production areas in six states of Brazil (Table 1). Symptomatic canes were obtained from selection areas managed by "Instituto Agronômico de Campinas (IAC) - Centro de Cana", located in Ribeirão Preto, Brazil. Six whips were obtained from three different sub-stations of the "Agro-industrial Experimental Station Obispo Colombres (EEAOC)" in Tucumán state (TUC), Argentina (Table 1). The reference SSC39A and SSC39B haploid strains were submitted to the culture collection of the "Fundação Oswaldo Cruz – FIOCRUZ" under the accession numbers INCQS 40413 and INCQS 40412, respectively. The SSC39B strain,

whose complete genome was sequenced (assembly accession: ASM101084v1), is herein referred to as BR_39B.

Table 1. *S. scitamineum* isolates collected from symptomatic sugarcane plants and molecular strategies applied.

Isolate	Mating type	Location [City (State)]	Collection date	Sugarcane cultivar	AFLP	tel-RFLP	ITS	Approximate geographical coordinates
Brazilian strains								
BR_04	A	Barreiras (BA)	2009	IAC 91 5155	✓	✓	✓	12º 09' 10" S
	B							44º 59' 24" W
BR_05	A	Frutal (MG)	2009	SP791011	✓	✓	✓	20º 01' 29" S
	B							48º 56' 26" W
BR_11	A	Jaboticabal (SP)	2009	IAC911099	✓	✓		21º 15' 17" S
	B							48º 19' 20" W
BR_17	A	Macatuba (SP)	2009	CTC12	✓	✓		22º 30' 08" S
	B							48º 42' 41" W
BR_18	A	Paranaíba (MS)	2009	SP891115		✓	✓	19º 40' 38" S
	B							51º 11' 27" W
BR_24	B	Paranaíba (MS)	2009	IACSP944004		✓		19º 40' 38" S
								51º 11' 27" W
BR_31	A	Serra Azul (SP)	2009	IAC044084	✓	✓		21º 18' 39" S
	B							47º 33' 56" W
BR_33	B	Vista Alegre do Alto (SP)	2009	RB855453		✓		21º 10' 14" S
								48º 37' 45" W
BR_35	A	Limeira D´Oeste (MG)	2009	IACSP933046	✓	✓		19º 33' 04" S
	B							50º 34' 50" W
BR_36	A	Valparaíso (SP)	2009	RB72454	✓	✓		21º 13' 40" S
	B							50º 52' 06" W
BR_38	A	Guaíra (SP)	2009	SP903414	✓	✓		20º 19' 06" S
	B							48º 18' 38" W
BR_39	A	Ribeirão Preto (SP)	2009	IAC982053	✓	✓	✓	21º 10' 39" S
	B							47º 48' 37" W
BR_40	A	Araçatuba (SP)	2009	nr*	✓	✓		21º 12' 32" S
	B							50º 25' 58" W
BR_41	A	Conchal (SP)	2009	nr*	✓	✓		22º 19' 49" S
	B							47º 10' 21" W
BR_43	A	Paranaíba (MS)	2010	IACSP942094		✓		19º 40' 38" S
								51º 11' 27" W
BR_46	A	Paranaíba (MS)	2010	CTC15		✓		19º 40' 38" S
	B							51º 11' 27" W
BR_58	A	Goianésia (GO)	2009	IACSP044004		✓	✓	15º 19' 03" S
	B							49º 07' 03" W
BR_65	A	São José da Laje (AL)	2010	SP716849		✓	✓	09º 00' 35" S
	B							36º 03' 30" W

BR_74	A	São José da Laje (AL)	2010	"havaiana"		✓		09º 00' 35" S
	B							36º 03' 30" W
BR_85	A	Paranaíba (MS)	2010	RB855002		✓	✓	19º 40' 38" S
	B							51º 11' 27" W
BR_89	A	São Simão (SP)	2011	nr*		✓		21º 28' 45" S
	B							47º 33' 03" W
BR_90	A	São Simão (SP)	2011	nr*		✓		21º 28' 45" S
	B							47º 33' 03" W
Argentine strains								
AR_01	A	Cerco Represa (TUC)	2014	TUC 06-12		✓	✓	26º48' 30"S
	B							65º13' 33" W
AR_02	A	Cerco Represa (TUC)	2014	TUC 06-25		✓	✓	26º 48' 29"S
	B							65º 13' 35"W
AR_05	A	Ingenio Santa Ana (TUC)	2014	TUC 06-04		✓	✓	27º 28' 19"S
	B							65º 40' 53"W
AR_06	A	Ingenio Santa Ana (TUC)	2014	TUC 05-10		✓	✓	27º 28' 30"S
	B							65º 40' 55"W
AR_07	A	Ranchillos (TUC)	2014	CP 65-357		✓	✓	26º 56' 58"S
	B							65º 03' 00"W
AR_13	A	Ingenio Santa Ana (TUC)	2014	TUC 06-25		✓	✓	27º 28' 20"S
	B							65º 40 '50"W

*nr=not registered

2.2) *S. scitamineum* strains isolation

Whips and teliospores were collected and haploid strains with opposite mating-types were derived from each isolate. Shortly, teliospores were scraped off from the whips and dried during four hours at 37ºC. A pinch of teliospores was dipped in 1 mL of saline solution (NaCl 0.85M) and treated with 0.5 mg L^{-1} of streptomycin sulphate (Sigma) for 30 minutes. Serial dilutions were prepared, plated on YM medium (0.3% yeast extract, 0.3% malt extract, 0.5% soybean peptone, 1% D-glucose, 1.5% agar), and incubated in the dark for 24 h at 28 ºC. A single germinated spore was inoculated in YM liquid, incubated overnight at 28 ºC in orbital shaker, and then plated on solid YM to obtain single haploid sporidial colonies (strains). The resulting yeast-like cells were stained with 1% orcein solution to confirm single nuclei state. Strains derived from each teliospore were randomly selected for the plate mating assay. Because the genomic context of the two alleles controlling mating-types were determined for the *S. scitamineum* strains SSC39A (MAT2 allele) and SSC39B (MAT1 allele) (Taniguti *et al*., 2015), they were used as reference to assign strains to the A or B

mating types. The strains were grown in liquid YM medium in orbital shaker for 24 h at 28ºC. A droplet of each paired culture was mixed and placed on YM plates, and incubated overnight at 28 ºC. Successful matings were identified by the formation of white and fuzzy colonies typical of the filamentous growth. Total DNA was extracted from strains of both mating-types and *S. scitamineum* identity was confirmed by PCR amplification with modified primers bE4 and bE8 (Taniguti *et al*., 2015).

2.3) Molecular markers techniques

2.3.1)telRFLP. Total genomic DNA were extracted from haploid strains and completely digested with three restriction enzymes (*Eco*RI, *Hind*III, and *Pst*I) in independent assays. The digested DNA was probed with the insert of pTEL13 recombinant plasmid, which corresponds to fungal telomeric region (Levis *et al*., 1997). The plasmid was digested and the insert gel purified with "illustra™ GFX™ PCR DNA and Gel Band Purification kit" (GE Healthcare). Labelling and detection of telRFLP fragments were performed using AlkaPhos Direct kit (GE Healthcare) following the manufacturer's instructions. Only reproducible and not ambiguous bands were scored. *In silico* predictions of endmost chromosome fragment size were performed based on the complete genome sequence of *S. scitamineum* (Taniguti *et al*., 2015).

2.3.2) AFLP. Templates were prepared by digesting the total DNA with *Mse*I and either *Pst*I or *Eco*RI, followed by ligation of adaptors and selective PCR amplifications, according to Palhares *et al*. (2012). AFLP bands were resolved in denaturing polyacrylamide gels using the electrophoresis apparatus Sequi-Gen GT (Bio-Rad), and visualized by silver staining. All samples were analyzed in duplicates and only reproducible bands were scored. The 125 bp and 10 bp ladders (Invitrogen) were used in order to estimate the size of AFLP fragments. Additionally, polymorphic AFLP bands were excised directly from the gels, re-amplified, size fragmented in agarose gel, purified with "™ illustra GFX ™ PCR DNA and Gel Band Purification kit" (GE Healthcare) and sequenced in an ABI PRISM 3730 DNA Analyzer (Applied Biosystems).

2.3.3) ITS-nrDNA sequencing. The primers used to amplify the ITS region were designed by White *et al*. (1990) and modified by Stoll *et al*. (2003). The amplifications were performed with the high fidelity enzyme "KAPA HiFi HotStart DNA Polymerase" (Kapa

Biosystems) in a "Veriti® 96-Well" thermocycler (Applied Biosystems). PCR products were purified with "™ illustra GFX ™ PCR DNA and Gel Band Purification kit" (GE Healthcare) and sequenced in a ABI PRISM 3730 DNA Analyzer (Applied Biosystems).

2.4) Stability of telRFLP markers across generations

Simulations of selfing (allowing natural crossing between teliospores-derived cells of BR_39 isolate) and outcrossing (using controlled crossing between BR_18A and BR_39B strains) were made across two consecutive generations in greenhouse (Fig. 2). Single-budded sugarcane setts were inoculated with *S. scitamineum* by punctuation. For the selfing experiment, BR_39 teliospores were inoculated to produce the first inbreeding generation of progeny BR_39-F1. The resulting teliospores were separated into opposite mating-type strains and used to produce the next inbreeding generation (BR_39-F2).

For the outcrossing simulation, the strains BR_18A (from whip BR_18) and BR_39B (from whip BR_39) were chosen because they presented contrasting telRFLP profiles and came from different locations (Table 1). A mixture of the two strains was inoculated to produce the progeny BR_18x39-F1. The following progenies BR_18x39-F2 were produced by inoculating two sexually compatible sporidial colonies isolated from the whip of previous disease cycle.

At each generation, five strains of opposite mating-types had their telRFLP profiles obtained with the restriction enzyme *Eco*RI.

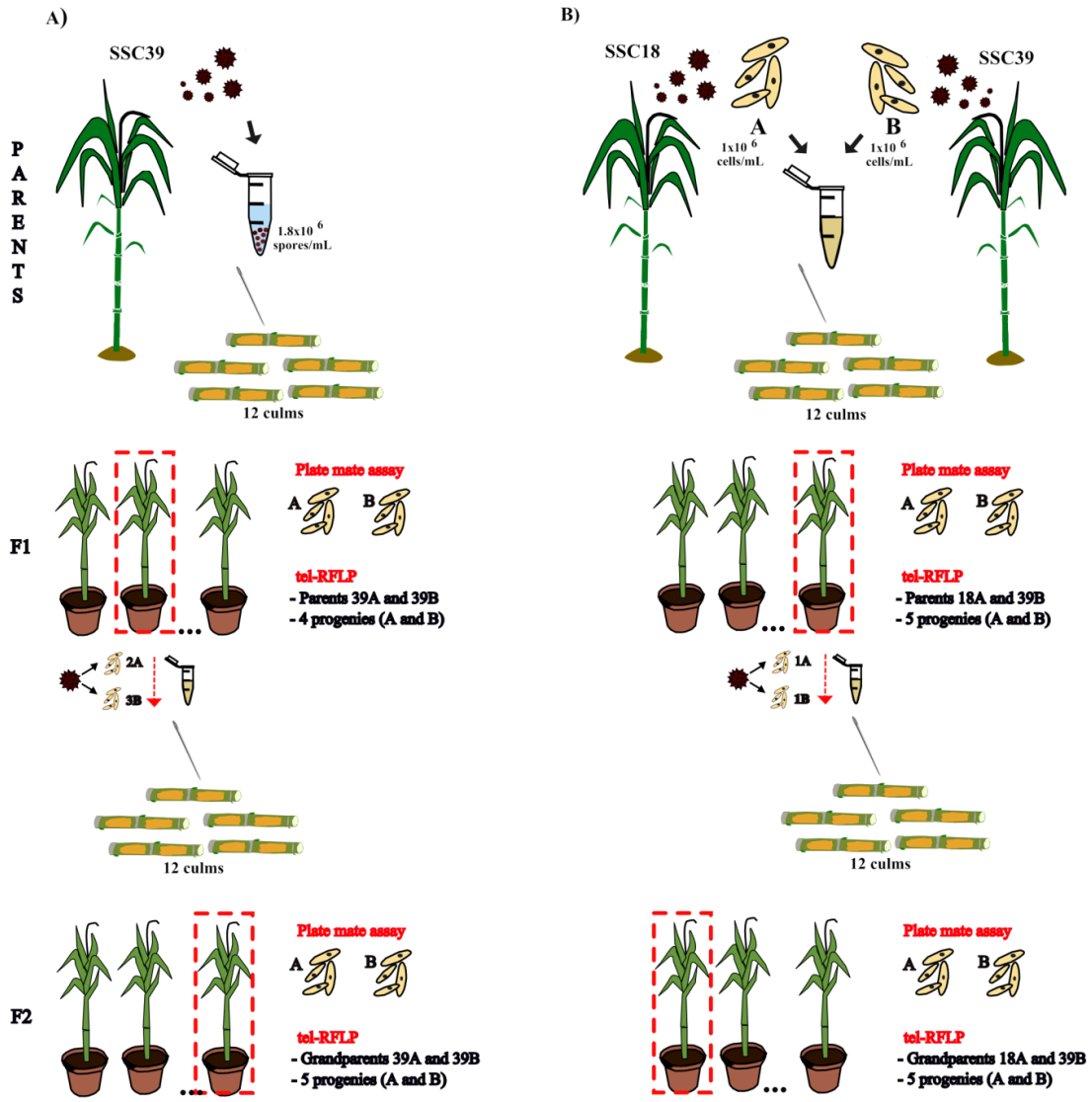

Fig. 2: Crossings conducted to evaluate the stability of *Eco*RI-telRFLP marks over generations. **A)** Selfing simulation between BR_39 teliospores. **B)** Outcrossing simulation between BR_18A and BR_39B yeast-like cells. At every generation, one whip was randomly chosen (plant surrounded by a red dotted line) to represent the filial generation through haploid strains isolation (A and B) and telRFLP profile determination. For each generation, six non-inoculated plants were used as control to ensure that the progenies were resulted from the artificial inoculations and not from contaminations.

2.5) Data analyses

Binary matrixes obtained from scoring telRFLP and AFLP polymorphic fragments as absent (0) or present (1) were used to estimate the genetic distance among *S. scitamineum* haploid strains using the Jaccard coefficient. When a marker was not scored for a particular strain, it was assigned as missing data. The distance matrix was used to construct a dendrogram based on the *Unweighted Pair Cluster Method with Arithmetic Mean* (UPGMA)

with 10 000 replications using the Pvclust package implemented in R (Suzuki & Shimodaira, 2006). The optimal number of genetic clusters (K) was estimated using the STRUCTURE v2.3 software (Pritchard *et al*., 2000), working with 125 000 burn-in steps before 500 000 MCMC repeats, for K ranging from 1 to 7 with 10 repeat runs for each K, under a model assuming admixture. The web-based program STRUCTURE HARVESTER was used to summarize the output data and to execute the "Evanno" method (Evanno *et al*., 2005), providing the K number that best fitted the data (Earl & vonHoldt, 2012). The Principal Component Analysis (PCA) for grouping the strains in the first two principal components (PCs) was conducted with the bioconductor package pcaMethods (Stacklies *et al*., 2007).

Mantel tests were performed with the "APE" package in R (Paradis *et al*., 2004) to verify the correlation between pairs of distance matrices (genetic, geographic, temporal, and breeding program). To generate a distribution of Z (Mantel's test statistic *Z*) under the null hypothesis of no association between matrices, 10 000 Monte Carlo random permutations were done. A two-tailed test was used by default for computing the statistical significance.

Sequencing data were trimmed and multiple-aligned using the genome sequence of SSC39B strain as reference (Taniguti *et al*., 2015) in the CLC Genomics Workbench v.8.0. As a single haplotype was obtained for ITS region in Brazilian and Argentine strains, the sequence can be recovered of the SSC39B genome sequence (accession assembly: ASM101084v1). Thirty ITS sequences from worldwide collected isolates were retrieved from Genbank (Table 2). A haplotype network was constructed based on ITS polymorphic sites using the Median-Joining method in the NETWORK v. 4.6.1.3 software (Fluxus Technology Ltd.).

Table 2. Internal Transcribed Spacer (ITS) sequences obtained from worldwide collected isolates of *S. scitamineum* retrieved from genbank.

Isolate	Mating type	Location [City (State)]	Collection date	Sugarcane cultivar	Acession Number	Source
Chinese isolates						
CH_01	H	Fujian	na*	Badila	EF185078	Zhang *et al*. (2015)
CH_02	H	Fujian	na*	Co1001	EF185077	Zhang *et al*. (2015)
CH_03	H	Fujian	na*	F134	EF185066	Zhang *et al*. (2015)
CH_04	H	Fujian	na*	NCo310	EF185069	Zhang *et al*. (2015)
CH_05	H	Fujian	na*	NCo376	EU427308	Zhang *et al*. (2015)
CH_06	H	Fujian	na*	Mingtang76-2	EF185076	Zhang *et al*. (2015)
CH_07	H	Fujian	na*	Guitang94-119	EF185073	Zhang *et al*. (2015)
CH_08	H	Jiangxi	na*	Guitang94-119	EF185071	Zhang *et al*. (2015)
CH_09	H	Jiangxi	na*	Mintang95-354	EF185081	Zhang *et al*. (2015)
CH_11	H	Guangdong	na*	ROC16	EF185082	Zhang *et al*. (2015)
CH_14	H	Hainan	na*	Guitang16	EF185075	Zhang *et al*. (2015)
CH_15	H	Hainan	na*	Yuetang92-126	EF185072	Zhang *et al*. (2015)
CH_16	H	Yunnan	na*	CP34-85	EF185074	Zhang *et al*. (2015)
CH_17	H	Yunnan	na*	Chuanzhe2	EF185068	Zhang *et al*. (2015)
CH_23	H	Yunnan	na*	ROC22	EU427309	Zhang *et al*. (2015)
Others						
ZA_01	A	South Africa	na*	na*	DQ004829	Singh *et al*. (2005)
ZA_01	B	South Africa	na*	na*	DQ004830	Singh *et al*. (2005)
ZA_02	H	South Africa	na*	na*	JN367296	Kellner *et al*. (2011)
RE_01	H	Reunion Island	na*	na*	DQ004832	Singh *et al*. (2005)
US_01	H	Hawaii - USA	na*	na*	DQ004833	Singh *et al*. (2005)
EG_01	H	Egypt	2008	GT54-9	JQ912112	Unpublished - NCBI
IN_01	H	India	na*	Co 96007	KP893340	Unpublished - NCBI
IN_03	H	India	2008	*S. officinarum*	FM179316	Unpublished - NCBI
TW_01	H	Taiwan	2010	ROC22	KJ467044	Unpublished - NCBI
TW_09	H	Taiwan	2010	ROC10	KJ194461	Unpublished - NCBI
TW_37	H	Taiwan	2011	ROC16	KJ467046	Unpublished - NCBI
JP_01	H	Japan	2008	*S.officinarum*	AB704871	Morita *et al*. (2012)
JP_02	H	Japan	2008	*S.officinarum*	AB704872	Morita *et al*. (2012)
JP_03	H	Japan	2008	*S.officinarum*	AB704873	Morita *et al*. (2012)
JP_04	H	Japan	2008	*S.officinarum*	AB704874	Morita *et al*. (2012)

na* = not available
H = hyphae

3) RESULTS AND DISCUSSION

3.1) Molecular variability of *S. scitamineum* unraveled by AFLP and telRFLP

In order to drawn a panoramic picture of the genetic variation of Brazilian isolates, we chose a combination of molecular approaches (telRFLP, AFLP, and ITS sequencing) to screen various regions of the genome and to compare markers used in previous studies.

Amplification with 19 AFLP combinations of restriction enzymes (*Eco*RI, *Pst*I and *Mse*I) and selective primers produced 36 polymorphic bands out of 1311 that were scored (Table 3). The number of polymorphic bands was in general low per gel (up to 3) and the highest number obtained was six with [P+AGA, M+CTC] and [P+AGA, M+CAA] combinations. This approach increased the number of polymorphic loci compared to the previous data for American isolates (Braithwaite *et al*., 2004), but still reveal limited variation among Brazilian strains. The ITS sequencing data was even more inefficient, producing identical sequences for all strains, including those from Argentina. The ITS data produced by other authors revealed some variability in isolates from different countries (Singh *et al*., 2005; Zhang *et al*., 2015). After align our data with other 30 ITS sequences available at *genbank*, the Brazilian and Argentine haplotypes were equal to the most common haplotype worldwide (Fig. 3).

Lastly, the telomeric hybridization patterns of genomic DNA digested with *Eco*RI, *Hin*dIII and *Pst*I resulted in the scoring of 18, 9 and 14 unambiguous polymorphic bands, respectively. Through telRFLP technique, almost a unique fingerprint was detected for each strain, unravelling a variation yet unknown for Brazilian isolates (Table 3).

Table 3 Variability detected by telRFLP and AFLP markers among sets of *S. scitamineum* strains.

Molecular Marker	Marker ID	Polymorphic bands	Strains	Haplotypes
telRFLP	*Eco*R1, *Hin*dIII, *Pst*I	41	53	45
	*Eco*RI	18	52	27
	*Hin*dIII	14	52	27
	*Pst*I	9	33	17
AFLP	*Eco*RI, *Pst*I, *Mse*I	36	24	18
Total	**all markers**	**77**	**53**	**45**

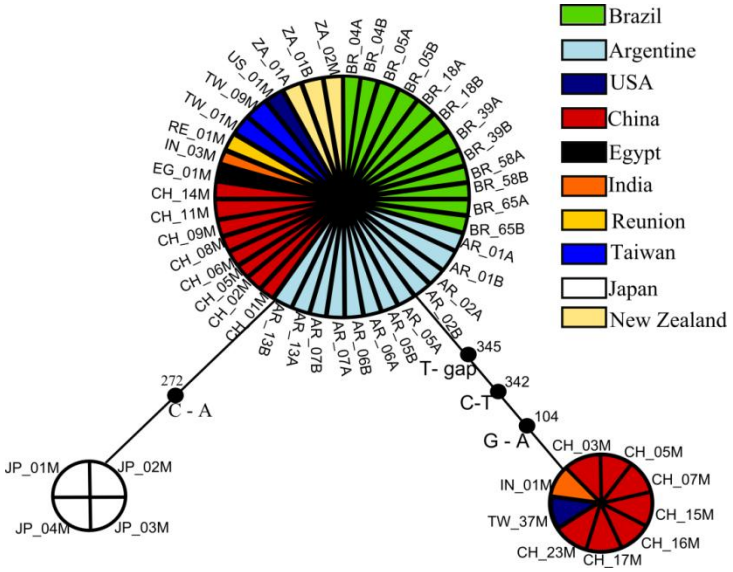

Fig. 3. Median Joining haplotype network from ITS sequences of *Sporisorium scitamineum* isolates (Table 2), obtained with the Network v. 4.6.1.3 software (Fluxus Technology Ltd.). Four parsimoniously informative sites were found in a total alignment of 581 bp. Each circle represents a haplotype. Colors indicate the different countries from where an isolate was collected. Dots in the branches connecting the circles show the mutational steps (type and position in the alignment) that differentiate the haplotypes. Brazilian and Argentine strains shared the same haplotype with isolates from several locations throughout the world.

From the 36 polymorphic AFLP fragments, only six could be recovered after reamplification and gel excision. The sequences were aligned by blastN against the SSC39B genome and the lowest e-values were scored in locations on chromosomes 02, 05, 08, 20, 22 and mtDNA (Fig. 4), which seems to be a random scanning of the genome.

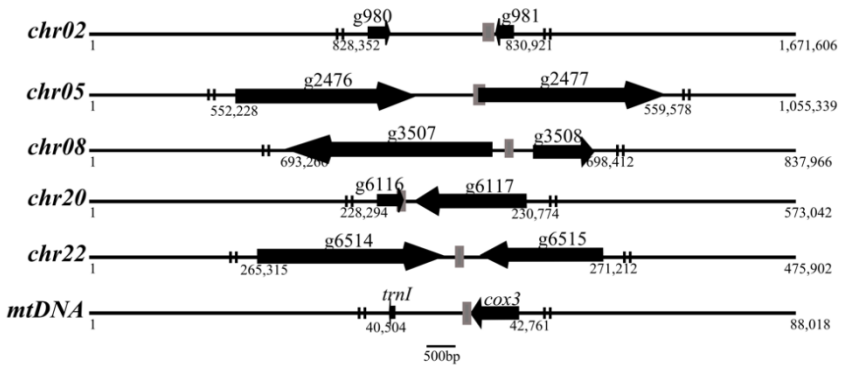

Fig.4: Genomic context of polymorphic bands retrieved from AFLP gels. Grey blocks indicate the position of the polymorphic bands sequenced in relation to SSC39B chromosomes, considering the lowest *e-value* by blastn alignment. Black arrows represent the genes surrounding each polymorphism pointing to transcription direction. Genes identification is the same as the automatically annotated by Taniguti *et al*. (2015). Numbers below the lines indicate the position on the chromosome. The genomic context is drawn to scale, but the total size of the chromosomes (from double bars) is out of scale.

Examining the *in silico* prediction of telRFLP endmost cutting sites in the genome of SSC39B (Taniguti *et al*., 2015), a list of target putative genes were detected (data not shown). Most of them were predicted members of the telomere-linked helicase (TLH) family that acts in telomere maintenance via recombination. Highly polymorphic TLHs are a common feature among filamentous fungi and yeast (Yamada *et al*., 1998; Gao *et al*., 2002; Sánchez-Alonso & Guzmán, 1998). Rehmeyer *et al*. (2009) have hypothesized that repeat-induced point mutation (RIP) could be the mutational process that generates the variation among multicopies of TLHs. Besides RIP mutations, a general source of telRFLP polymorphisms can be subtelomeric rearrangements during mitosis or meiosis that lead to chromosome end variability by duplication, deletion or (non)homologous recombination events (Louis & Haber, 1990; Winzeler *et al*., 2003; Zolan, 1995). The source of telRFLP variation in *S. scitamineum* may be generated by a combination of all these processes, and it seems to provide a better source of genetic variation among strains.

3.2) Relationship among Brazilian and Argentine strains

The combined data analysis of the two markers was conducted totaling 77 polymorphic *loci*. Two main groups were identified by the UPGMA dendrogram (Fig. 5A). A and B strains isolated from the same diseased plant were clustered into the same main group, except for BR_41 isolate (Fig. 5A). The prevalence of selfing explains the clustering pattern of A and B strains (Raboin *et al*., 2007). However, these strains did not have identical markers profile, evidencing the presence of variation even between strains de rived from the same teliospore.

A model-based clustering algorithm, performed with STRUCTURE software, identified that the most likely number of clusters is three (K=3) and provided the quantification of membership to each cluster for each strain (Fig. 5B). Most individuals (42) have very high membership coefficients to a single cluster (>=0.9), but an admixture composition was noticed for some strains, whose genetic makeup is drawn from more than one of the *K* clusters. The STRUCTURE results arranged by line according to the dendrogram were not corroborative. When the dendrogram was partitioned into three groups (which is also well supported), the STRUCTURE red cluster fully covered one dendrogram group and partially another. The Principal Components Analysis (PCA) was used as a third method to examine the clustering pattern of *S. scitamineum* strains, producing a two-dimensional graphic summarizing the maximal variance of the multi*locus* data (Fig. 5C). The

three clusters generated by the STRUCTURE software were not consistent with the PCA results. The two groups set up by the dendogram can be viewed within ellipses.

The whip-like structure harboring the teliospores emerges two months after inoculation in highly susceptible varieties (Legaz *et al*., 2011). Spores are carried by the wind over a distance of 100 meters from the source whip and by splash of rainfall water in the soil (Hoy *et al*., 1991; Sreeramulu & Vittal, 1972). Despite the potential scattering over long distances, we found no association between genetic and geographic distances among isolates (Fig. 5D and 5E), which was confirmed by the Mantel correlation test (Fig. 6). The genetic distance matrix among *S. scitamineum* strains were not significantly correlated with any variable tested (spatial, temporal and breeding program origin), whereas these other variables were correlated between them. The dispersion within the Brazilian territory and among countries was probably due to human activity by exchanging asymptomatic plant material, as already proposed on global scale by Braithwaite *et al*. (2004) and Raboin *et al*., (2007).

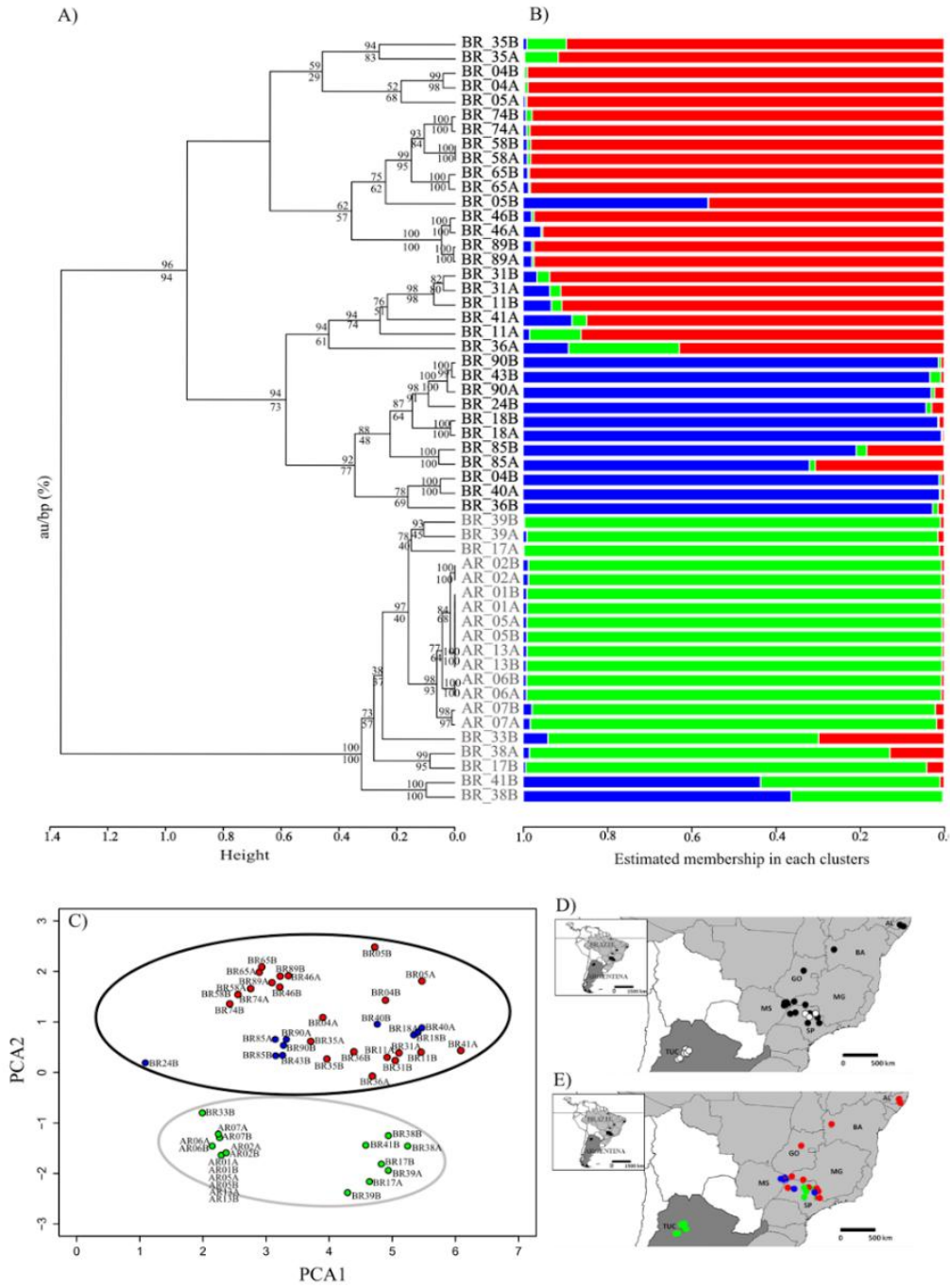

Fig. 5. Relationship among *Sporisorium scitamineum* strains using telRFLP and AFLP markers. **A)** UPGMA dendrogram of genetic distance using Pvclust R package. Black and grey colors were attributed to labels according to the two groups defined. Values at branches are AU p-values (above) and BP values (below). **B)** Data structure obtained by STRUCTURE software for the best estimated value of K clusters (K=3). Each strain is represented by a horizontal bar, ordered according to the dendrogram. Each bar is partitioned into three colored segments (red, blue and green) that represent the estimated membership in each cluster. **C)** Graphic representation of the first two principal components (PCAs). Each point represents a strain. The color of the points corresponds to the primary color assigned by STRUCTURE analysis. The two ellipses indicate the groups generated by UPGMA dendrogram. **D)** and **E)** Approximate geographic distribution of *S. scitamineum* isolates (colored dots) included in this study. Dot colors indicate the grouping pattern generated by the dendrogram (A) and STRUCTURE (B) analysis, respectively.

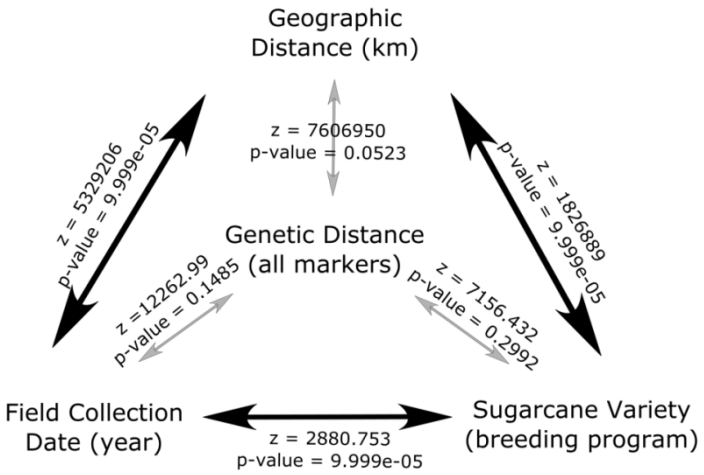

Fig. 6. Mantel correlation tests between four distance matrices (genetic, geographic, temporal, and breeding program) using the APE R package. Black arrows represent statistically significant associations by two-side Z-statistics test, while gray arrows represent no significant associations.

3.3) telRFLP marker over generations

Sporisorium scitamineum is a dimorphic fungus that requires sexual reproduction to switch from a saprophytic budding yeast phase to an infectious dikaryotic hyphal growth. This transition is essential for the completion of its life-cycle, since sporulation is host-dependent (Bakkeren *et al*., 2008). As an obligate sexual fungus that requires different mating-types in the haploid strains for successful sexual reaction, one would expect high genotypic diversity in its population. The admix of pre-existing genetic variation due to sexual recombination can be observed in our outcrossing simulation experiment between *S. scitamineum* strains (BR_18A x BR39_B) in which occurs the reassortment of parental telRFLP polymorphisms into the progenies. However, *S. scitamineum* has a bipolar mating system and selfing seems to be the predominant sexual reproduction mode in natural populations (Raboin *et al*., 2007).

The parental strains derived from the BR_39 whip had identical telRFLP *Eco*RI profiles, whereas the parental strains BR_18A and BR_39B differed at four scored bands. Interestingly, a different hybridization pattern emerged in F1 progenies derived from selfing; whereas in the outcrossing simulation no new pattern was detected (data not shown).

4) CONCLUSION

This study provides new clues about informative regions of the *S. scitamineum* genome, revealing genetic variability among Brazilian strains that so far were not described. No polymorphism was found by using ITS sequences, but taken together AFLP and telRFLP

markers generate almost a unique fingerprint for each strain. The fifty-three strains were clustered into two genetically distinct major groups that do not reflect their geographical origins. We suggest that a mixture of isolates from these two major genetic backgrounds should be used in breeding programs to cover different adaptive potentials of Brazilian isolates. Our results do not allow us to infer about the emergence of new races, but showed that there is variability among strains and that new genetic variants may appear in every generation. It is worth mentioning that population genomics studies of this pathogen can also bring new knowledge to light regarding other polymorphic genomic regions and for the understanding of the disease dispersion history.

5) JOURNAL ACKNOWLEDGMENT

This is a pre-copyedited, author-produced version of an article accepted for publication in FEMS Microbiology Letters following peer review. The version of record "Benevenuto J. *et al*. (2016). Molecular variability and genetic relationship among Brazilian strains of the sugarcane smut fungus. FEMS Microbiology Letters 363(24)" is available online at: DOI: https://doi.org/10.1093/femsle/fnw277.

REFERENCES

Alexander K & Srinivasan K (1966). Sexuality in Ustilagoscitaminea Syd. *Current Science* **35**: 603–604.

Askree SH, Yehuda T, Smolikov S *et al.* (2004). A genome-wide screen for *Saccharomyces cerevisiae* deletion mutants that affect telomere length. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 8658–8663.

Bakkeren G, Kämper J & Schirawski J (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. *Fungal Genetics and Biology* **45**: 15-21.

Bhuiyan SA, Croft BJ, Stringer JK & Deomano EC (2015). Pathogenic variation in spore populations of Sporisorium scitamineum, causal agent of sugarcane smut in Australia. *Plant Disease* **99**:93-99.

Billiard S, López-Villavicencio M, Hood ME & Giraud T (2012). Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *Journal of Evolutionary Biology* **25**: 1020–1038.

Braithwaite KS, Bakkeren G, Croft BJ & Brumbley SM (2004). Genetic variation in a worldwide collection of the sugarcane smut fungus *Ustilago scitaminea*. *Proceedings of the Australian Society of Sugar Cane Technology* **26:** 1–9.

Comstock J (2000). Smut. In: Rott P, Bailey R, Comstock J, Croft B & Saumtally, A. (eds.), *A Guide to Sugarcane Diseases*. CIRAD - ISSCT, Montpellier, pp. 181–185.

Comstock J & Lentini R (2005). Sugarcane smut disease. SS-AGR-208, Florida Cooperative Extension Service, Institute of Food Agricultural Sciences, University of Florida.

Earl D & vonHoldt B (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

Evanno G, Regnaut S & Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–20.

Fan HY & Klein HL (1994). Characterization of mutations that suppress the temperature-sensitive growth of the *hpr1* delta mutant of *Saccharomyces cerevisiae*. *Genetics* **137**: 945–956.

Fattah AE, Alamiri S, Abou-Shanab R & Hafez EE (2010). Fingerprinting of *Ustilago scitaminea* (Sydow) in Egypt using differential display technique : chitinase gene the main marker. *Journal of Agricultural and Biological Science* **6:** 8–13.

Ferreira SA & Comstock JC (1989). Smut. In: Ricaud C, Egan BT, Gillaspie AG & Hughes CG. (eds.), *Diseases of Sugarcane*. Elsevier, Amsterdam, pp. 211–229.

Fraser JA & Heitman J (2003). Fungal mating-type *loci*. *Current Biology* **13**: 792–795.

Gao W, Khang CH, Park S-Y, Lee Y-H & Kang S (2002). Evolution and organization of a highly dynamic, subtelomeric helicase gene family in the rice blast fungus *Magnaporthe grisea*. *Genetics* **162**: 103–112.

Giraud T, Yockteng R, López-Villavicencio M, Refrégier G & Hood ME (2008). Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. *Eukaryotic Cell* **7**: 765–775.

Hoy J, Grisham M & Chao C (1991). Production of sori and dispersal of teliospores of *Ustilago scitaminea* in Louisiana. *Phytopathology* **81**: 574–579.

Kaltz O & Shykoff JA (1999). Selfing versus outcrossing propensity of the fungal pathogen *Microbotryum violaceum* across *Silene latifolia* host plants. *Journal of Evolutionary Biology* **12**: 340–349.

Legaz ME, Santiago R, Armas R *et al*. (2011). Molecular defence responses of sugarcane (*Saccharum officinarum L*.) to smut (*Sporisorium scitamineum* (Syd.) Piepenbr&Oberw, 2002). In Méndez-Villas A (ed), *Science against microbial pathogens: communicating current research and technological advances*, 3rd edn. Formatex, Badajoz, pp. 1244–1250.

Levis C, Giraud T, Dutertre M, Fortini D & Brygoo Y (1997). Telomeric DNA of *Botrytis cinerea*: a useful tool for strain identification. *FEMS Microbiology Letters* **157**: 267–272.

Louis EJ & Haber JE (1990). Mitotic recombination among subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics* **124**: 547–559.

Luzaran RT, Cueva FMD, Cumagun CJR, Velasco LRI & Dalisay TU (2012). Variability of sugarcane smut pathogen, *Ustilago scitaminea* Sydow in the Philippines. *Philippine Journal of Crop Science* **37**: 38–51.

Mueller UG & Wolfenbarger LL (1999). AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14**: 389-394.

Nilsson HR, Kristiansson E, Ryberg M, Hallenberg N, Larsson K-H (2008). Intraspecific *ITS* Variability in the Kingdom *Fungi* as Expressed in the International Sequence Databases and Its Implications for Molecular Species Identification. Evol Bioinform Online, **4**: 193–201.

Palhares AC, Rodrigues-Morais TB, Van Sluys MA *et al*. (2012). A novel linkage map of sugarcane with evidence for clustering of retrotransposon-based markers. *BMC Genetics* **13**: 1-16.

Paradis E, Claude J & Strimmer K (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.

Pritchard JK, Stephens M & Donnelly P (2000). Inference of population structure using multi*locus* genotype data. *Genetics* **155**: 945–959.

Que Y, Xu L, Lin J, Chen R & Grisham MP (2012). Molecular variation of Sporisorium scitamineum in Mainland China revealed by RAPD and SRAP markers. *Plant Disease* **96**:1519-1525.

Raboin L-M, Selvi A, Oliveira KM *et al*. (2007). Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen Ustilago scitaminea. Fungal Genetics and Biology **44**: 64–76.

Rago AM, Casagrande MV & Massola-Júnior, NS (2009). Variabilidade patogênica de Ustilago scitaminea no estado de São Paulo. Summa Phytopathologica, **35**: 93-97.

Rehmeyer C, Li W, Kusaba M & Farman M (2009). The telomere-linked helicase (TLH) gene family in *Magnaporthe oryzae*: revised gene structure reveals a novel TLH-specific protein motif. *Current Genetics* **55**: 253–262.

Sánchez-Alonso P & Guzmán P (1998). Organization of chromosome ends in *Ustilago maydis*: RecQ-like helicase motifs at telomeric regions. *Genetics* **148**: 1043–1054.

Singh N, Somai BM & Pillay D (2005). Molecular profiling demonstrates limited diversity amongst geographically separate strains of *Ustilago scitaminea*. *FEMS Microbiology Letters* **247**: 7–15.

Sreeramulu T & Vittal BPR (1972). Spore dispersal of the sugarcane smut (*Ustilago scitaminea*). *Transactions of the British Mycological Society* **58**: 301–312.

Stacklies W, Redestig H, Scholz M, Walther D & Selbig J (2007). pcaMethods a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**: 1164–1167.

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. *Canadian Journal of Botany* **81**: 976–984

Suzuki R & Shimodaira H (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540–1542.

Taniguti LM, Schaker PDC, Benevenuto J *et al.* (2015). Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with sugarcane. *PLoS One* **10**: e0129318.

White TJ, Bruns T, Lee S, Taylor J (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfund D, Sninsky J, White T (eds), *PCR Protocols: A Guide to Methods and Application*. Academic Press, New York, pp. 315–322.

Winzeler EA, Castillo-Davis CI, Oshiro G *et al.* (2003). Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics* **163**: 79–89.

Wu C, Kim Y-S, Smith KM, Li W, Hood HM, Staben C, Selker EU, Sachs MS & Farman ML (2009). Characterization of Chromosome Ends in the Filamentous Fungus Neurospora crassa. Genetics **181**: 1129–1145.

Xu L, Lu Y, You Q *et al.* (2014). Biogeographical variation and population genetic structure of *Sporisorium scitamineum* in Mainland China: Insights from ISSR and SP-SRAP markers. *The Scientific World Journal* **2014**:1-13.

Xu L, Que Y & Chen R (2004). Genetic diversity Mainland China of *Ustilago scitaminea*. *Sugarcane Biotechnology* **6**: 267–271.

Yamada M, Hayatsu N, Matsuura A & Ishikawa F (1998). Y′-Help1, a DNA Helicase encoded by the yeast subtelomeric Y′ element is induced in survivors defective for telomerase. *Journal of Biological Chemistry* **273**: 33360–33366.

Zhang YY, Huang N, Xiao XH *et al.* (2015). Molecular variation of *Sporisorium scitamineum* in Mainland China revealed by internal transcribed spacers. *Genetics and Molecular Research* **14**: 7894–7909.

Zolan ME (1995).Chromosome-length polymorphism in fungi. *Microbiological Reviews* **59**: 686–698.

CHAPTER 2: Polymorphic variant of a candidate effector potentially involved in the specific interaction between *Sporisorium scitamineum* and sugarcane

ABSTRACT

Sporisorium scitamineum is a biotrophic pathogen establishing a unique interaction with sugarcane. Secreted effectors proteins play a central role in biotrophic interactions, suppressing the plant's immune system and altering the host metabolism to support the fungal growth and development. Effectors are emerging as tools in resistance breeding programs to accelerate the identification of pathogen races and host resistance genes. Through computational predictions and transcriptomic analysis, a *S. scitamineum* putative secreted effector, unique to this fungus with regard to three closest species, and exclusively expressed during the sugarcane interaction in relation to culture medium growth, was identified. This gene, automatically named as *g1052* in the genome annotation, was used in the present study to perform a screening into 52 haploid strains of *S. scitamineum* derived from 28 diseased plants naturally found in different regions of Brazil and Argentina. After gene amplifications and further sequencing, two polymorphisms were found, one at position 25 and another at position 472 from a total of 513 base pairs alignment (comprising from the start to the stop codon). Each polymorphism affected the first base of the respective codon, leading to non-synonymous substitutions. Only two haplotypes were obtained. Maximum Likelihood tree based on the protein sequences showed that there is no variation among strains derived from the same diseased plant and Brazilian and Argentine strains shared one of the haplotypes. Additional experiments are needed to characterize the function of this gene and to associate the haplotypes with differences in aggressiveness that could indicate the presence of races.

Keywords: Ssugarcane smut; Effector; Non-synonymous polymorphism: Haplotypes

1) INTRODUCTION

Fungal plant pathogens are able to infect, colonize and uptake nutrients from their hosts by using effector proteins. Effectors are defined as pathogen secreted molecules that alter host-cell structure and function, thereby promoting the disease or triggering defense responses (Kamoun, 2006; Selin *et al*.,2016). Therefore, effectors can be virulence or avirulence factors during host interaction. Most effectors were first identified due to their avirulence activity (Avr genes), i.e., by their ability to activate hypersensitive cell death response (HR) when recognized by a plant resistance gene (R) (Lo Presti *et al*., 2015). The recent advances in high-throughput sequencing technologies allowed the genome-wide prediction of effector genes, opening new doors for the understanding of complex diseases.

Sporisorium scitamineum is a biotrophic pathogen establishing a unique interaction with sugarcane and causing smut disease. The *S. scitamineum* genome and transcriptome were obtained by Taniguti *et al*. (2015). The predicted effector repertoire comprised genes encoding 70 proteins with signal peptide for secretion, no trans-membrane domains, no

glycosylphosphatidylinositol (GPI) anchor, small size (<30 kDa) and high cysteine content (>1,5%). Some of those genes were differentially expressed at early (5 days after infection) and late (200 days after infection) moments of the disease in relation to the *in vitro* growth. Moreover, *S. scitamineum* specific effectors were also detected through orthologous analysis comparing to the close related species *S. reilianum, U. maydis* and *U. hordei* (Taniguti *et al*., 2015). Considering the computational prediction of effectors, the differential expression *in planta* and specificity to *S. scitamineum* genome, the candidate effector gene *g1052* was chosen for a variability screening among *S. scitamineum* isolates.

Due the key role exerted by effectors during plant-pathogen interactions, they are predicted to be under strong host selective pressures, whereby alternative forms that do not trigger defense will be favored. Effector genes are frequently polymorphic, present signatures of diversifying selection, and lack homologs in closely related species (Sperschneider *et al*., 2015). Monitoring of effector allele diversity in pathogen populations can assist in plant breeding programs though the early detection of races that can overcome the deployed plant resistance genes (Pais *et al*., 2013). Hence, this study aims to identify and characterize polymorphic sites at a candidate effector gene *g1052* in Brazilian and Argentine isolates of the fungus *S. scitamineum*.

2) MATERIAL AND METHODS

Fifty-three haploid strains derived from *S. scitamineum* isolates collected in twenty-two localities of Brazil and six localities of Argentina were used for effector variability screening. See Chapter 1-Table 1 for strains information and isolation.

Primers flanking the *g1052* gene were designed using the genome of *S. scitamineum* (SSC39B) as reference (Taniguti *et al*., 2015): primer F (5'- GATTGGATGGTGTGGATG -3') and primer R (5'-TGTGGGAGTAAAGATGGTAG -3'). Conventional PCR reactions were conducted using the high fidelity enzyme "KAPA HiFi Hot Start DNA Polymerase" (Kapa Biosystems), according to the manufacturer's instructions. Amplified fragments were purified with "illustra™ GFX™ PCR DNA and Gel Band Purification" kit (GE Healthcare) and sequenced using ABI Prism 3100 Genetic Analyser (Applied Biosystems).

Sequences were trimmed and aligned using CLC Genomics Workbench v.8 (CLC Bio). *g1052* sequences were also retrieved from Chinese and Australian strains, whose genome are available (GenBank assembly accession: GCA_000772675.1 and GCA_001243155.1, respectively). The DNA-coding sequences were translated to the predicted protein sequences. The Poisson's distance was used to construct a dendrogram based

on *Unweighted Pair Cluster Method with Arithmetic Mean* (UPGMA) with 100 replications using MEGA v.5 (Tamura *et al.*, 2011). *Phobius* web server (www.phobius.sbc.su.se) was used for signal peptide and transmembrane topology prediction.

3) RESULTS AND DISCUSSION

Two polymorphisms were found after gene alignment for the 55 strains analyzed. One polymorphism was located at position 25 and another at position 472 from a total of 513 base pairs alignment (comprising from the putative start to the stop codons). Each polymorphism affected the first base of the respective codons, leading to non-synonymous amino acid substitutions.

The first polymorphism occurred at the predicted peptide signal, specifically in the hydrophobic position (H-region) and causes an Alanine-to-Threonine (Ala to Thr) substitution (Figure 1). Both amino acids have small sizes and the substitution is considered neutral for many protein types. The second polymorphism occurred at the non-cytoplasmic portion of the protein and causes an Arginine-to-Glycin (Arg to Gly) substitution (Figure 1). The Arg-Gly substitution is disfavored in many protein types because they differ in size, polarity, and charge.

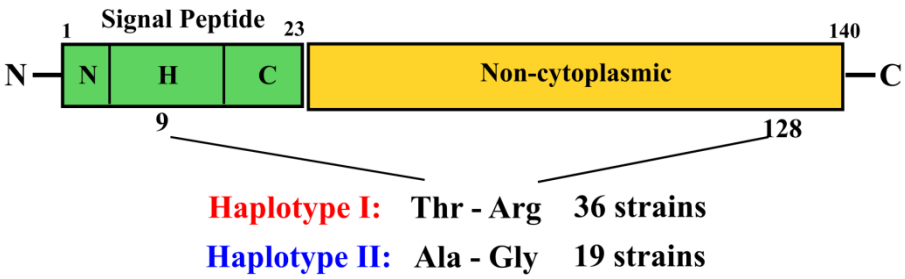

Figure 1. Schematic representation of G1052 protein and haplotypes. Green and yellow colors represent signal peptide and non-cytoplasmic portion, respectively. Numbers above the drawing show the size of the partitions defined by Phobius software and numbers below show the substitution sites. The two haplotypes (red and blue) and the number of strains harboring each haplotypes are indicated.

It is also noteworthy that only two haplotypes were found: Thr-Arg (haplotype I) and Ala-Gly (haplotype II). The haplotype I is shared by most of Brazilian strains and also by the Australian (AU_CCFA01) and Chinese (CH_JFOL01) strains. Haplotype I is shared between Argentine and some Brazilian strains (Figure 2). The distribution of the two haplotypes among Brazilian and Argentine strains reflects the two large groups formed by analysis with AFLP and telRFLP markers (see Chapter 1).

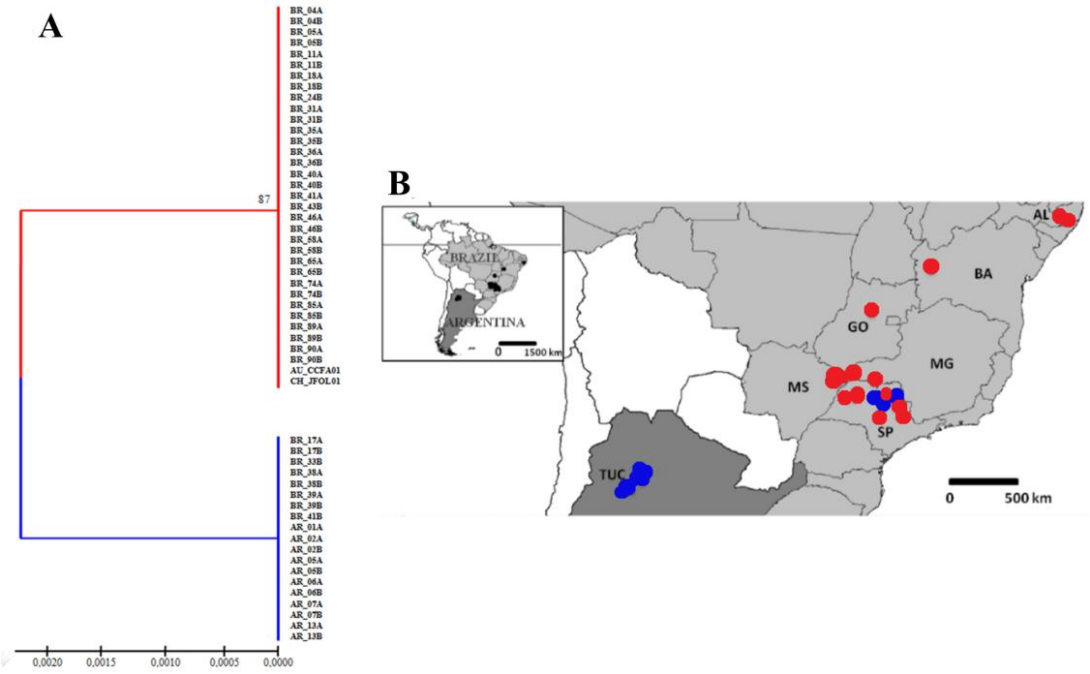

Figure 2. UPGMA dendrogram (A) and geographic distribution (B) of the two *g1052* haplotypes. Haplotypes I and II are represented as red and blue, respectively.

Variants of effector genes in populations of the rust pathogen *Melampsora line* were differentially recognized by resistance genes in host *Linun marginale* (Barret *et al*., 2009). Polymorphic effectors could also differentiate races of *Plasmopara halstedii* (the causal agent of downy mildew on sunflower) (Gascuel *et al*., 2016) and *Colletotrichum lentis* (the causal agent of anthracnose on lentil) (Bhadauria *et al*., 2015). Hence, important forward steps should be experimentally characterize the *g1052* candidate effector in S. *scitamineum*-sugarcane interaction and verify if the two polymorphic variants entail differences in recognition or in efficiency in infecting/colonizing sugarcane genotypes.

4) CONCLUSION

In this study, we identified the presence of two polymorphic sites in a candidate effector gene potentially involved in the specific interaction between *S. scitamineum* and sugarcane. Additional experiments are needed to characterize the function this gene and verify whether the two haplotypes of *g1052* imply variations in pathogen aggressiveness degree, which may indicate the presence of pathogenic races and can further assist in breeding programs.

REFERENCES

Barrett LG, Thrall PH, Dodds PN, van der Merwe M, Linde CC, Lawrence GJ, Burdon JJ (2009). Diversity and Evolution of Effector Loci in Natural Populations of the Plant Pathogen Melampsora lini. *Molecular Biology and Evolution* **26**(11): 2499–2513.

Bhadauria V, MacLachlan R, Pozniak C, Banniza S (2015). Candidate effectors contribute to race differentiation and virulence of the lentil anthracnose pathogen Colletotrichum lentis. *BMC Genomics* **16**(1): 628.

Gascuel Q, Bordat A, Sallet E, Pouilly N, Carrere S, Roux F, Vincourt P, Godiard L. (2016). Effector Polymorphisms of the Sunflower Downy Mildew Pathogen Plasmopara halstedii and Their Use to Identify Pathotypes from Field Isolates. PLoS One 11(2): e0148513.

Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual Review of Phytopathology* **44**:41-60.

Lo Presti L, Lanver D, Schweizer G, *et al*. (2015). Fungal effectors and plant susceptibility. *Annual Review of Plant Biology* **66**: 513–545.

Pais M, Win J, Yoshida K, Etherington G, Cano L, Raffaele S, Banfield M, Jones A, Kamoun S, and Saunders DGO (2013). From pathogen genomes to host plant processes: The power of plant parasitic oomycetes. *Genome Biology* **14** (6):211.

Selin C, Kievit TR, Belmonte MF, and Fernando, DWG (2016). Elucidating the role of effectors in plant-fungal interactions: progress and challenges. *Frontiers in Microbiology* **7**: 600.

Sperschneider J, Dodds PN, Gardiner DM, Manners JM, Singh KB, Taylor JM. (2015). Advances and challenges in computational prediction of effectors from plant pathogenic fungi. *PLoS Pathogens* **11**(5): e1004806.

Taniguti LM, Schaker PDC, Benevenuto J, *et al.* (2015). Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with sugarcane. *PLoS One* **10**: e0129318.

CHAPTER 3: Characterization of mating-type *loci* in *Sporisorium scitamineum* genome and implications in smut fungi evolution

ABSTRACT

Smut fungi need to undergo a successful mating reaction to form the infective dikaryotic hyphae. Two loci, named *a* and *b*, are involved in mating-type determination. Given the importance of mating-type loci for orchestrating the sexual cycle, establishment of cell-type identity, triggering pathogenicity, and influencing the evolutionary trajectory of smut pathogens, the two opposite mating-type loci and their flanking regions were characterized in *Sporisorium scitamineum* genome and compared with other available sequences from Ustilagomicotina fungi. The *a* and *b* loci are linked at 59 kbp apart on chromosome 2 of *S. scitamineum*, characterizing a bipolar mating system. Transposable elements are the likely mechanism for chromosomal rearrangements leading to linkage of *a* and *b* loci. Mating-type proteins have lower percentages of identity among smut species in comparison with the average of all predicted proteins. Alleles of *b* locus were more close related to each other than to alleles from others species; whereas proteins encoded by the *a* locus clustered preferentially with alleles from others species. The presence of trans-specific polymorphisms at *a* locus suggests that receptors of one species may recognize pheromone from another species and hybridization events may had happened during smut fungi evolution.

Keywords: Smut fungi; Mating-type; Bipolar; Transposable elements; Hybridization

1) INTRODUCTION

Smut fungi need to undergo a successful mating reaction to form infective dikaryotic hyphae. Both process of fusion between compatible haploid yeast-like cells and the triggered pathogenicity are regulated by two *loci*: *a* and *b* (Bakkeren *et al*., 2008). At the *a locus*, genes encode a pheromone/receptor system necessary for cell-cell recognition and fusion, while at the *b locus* there are two genes encoding subunits of a heterodimeric homeodomain transcription factor that is a key regulator for maintenance of the dikaryon as well as for pathogenicity. Only the combination of proteins at both *loci* from different mates is able to initiate sexual development and pathogenicity. Thereafter, the maintenance of allelic variants of mating-type genes is essential to sexual reproduction and is expected to be subject to balancing selection in the population (Kellner *et al*., 2011).

Differences in structure of *a* and *b loci* imply differences in mating behavior as tetra or bipolar system. In tetrapolar systems, *a* and *b loci* are unlinked and segregate independently, which means that from one spore four mating-types could be isolated (Schirawski *et al*., 2005). In bipolar systems, *a* and *b loci* are linked and segregate as one *locus*, i.e., after meiosis one spore could give rise to only two different mating-types sporidia cells. In general,

the bipolar system favors a higher level of selfing compared with the tetrapolar system, leading to elevated homozygosity in the fungal population (Fraser & Heitman, 2003), reducing the reassortment of genetic variability, and thus the potential of the pathogen to adapt to evolutionary changes of the host (Kaltz & Shykoff, 1999). On the other hand, selfing could be selectively advantageous in cases of limited partner availability, ensuring reproductive success (Billiard *et al*., 2012). The advantages of inbreeding or outbreeding may depend on the ecological niche occupied by the fungus (Bakkeren *et al*., 2008) and their occurrence is also dependent on the number and frequency of alternate alleles at the mating-type *loci* (Giraud *et al*., 2008).

Given the importance of mating-type *loci* for orchestrating the sexual cycle, establishment of cell-type identity, triggering pathogenicity, and influencing the evolutionary trajectory of smut pathogens, we characterized the two opposite mating-type *loci* in *S. scitamineum* and their flanking regions and compared with other available sequences from Ustilagomicotina fungi.

2) METHODOLOGY

2.1) Annotation of mating-type genes

S. scitamineum SSC39B and SSC39A strains are opposite mating-type haploid cells derived from a single teliospore (SSC39). The complete genome assembly was obtained for SSC39B strain and BAC genomic library sequences targeted to mating-type genes were obtained for SSC39A strain (Taniguti *et al*., 2015). The mating type genes at *a locus* (*pra* and *mfa*) and *b locus* (*bE* and *bW*) were manually annotated for both strains. BLASTx (Altschul *et al*., 1990) was used into CLC Genomics Workbench v.7.0 (CLC Bio) to find regions of local similarity, considering the *S. scitamineum* assemblies as a query and proteins from related species as subject. Artemis 16.0 (Rutherford *et al*., 2000) was used to correct the ORF of the sequences. We also observed the alignment of reads from RNAseq of the fungus (Taniguti *et al*., 2015) against these genes using CLC Genomics Workbench v.7.0.

In addition, the inter*loci* region in the SSC39B genome was screened for repetitive elements using RepeatMasker web server (http://www.repeatmasker.org).

2.2) Smut genomes comparison

The proteomes predicted from other smut genome assemblies (*S. reilianum*, *Ustilago maydis* and *U. hordei*) were compared with the *S. scitamineum* proteins in relation to the

percentage identity by BLASTp. Synteny between chromosomes of *S. reilianum* and *S. scitamineum* harboring the mating-type genes were performed using BLASTn (cutoff e-value ≤ 1 x 10-5) and visualized using Kablammo (http://kablammo.wasmuthlab.org).

The genomic context was designed for *S. scitamineum* and close related species using the length (base pairs) of each mating type genes and the distance between flanking genes.

2.5) Reconstruction of phylogenetic trees

Phylogenetic analysis of mating-type protein sequences were performed based on a multiple sequence alignment generated by T-COFFEE (Notredame *et al*., 2000). The best amino acids substitution model that fits the data was determined by using the Akaike Information Criterion (AIC) in the software ProtTest v3.2 (Darriba *et al*., 2011). Maximum likelihood trees were obtained for each protein considering the heuristic method NNI (Nearest Neighbor Interchange) for searching through treespace and aLRT SH-like (approximate Likelihood-Ratio Test with Shimodaira-Hasegawa-like procedure) for quantifying branch support using PhyML v3.0 (Guindon *et al*., 2010). *S. scitamineum* protein sequences of both mating-types and sequences for close related species available on GenBank were included in the analysis.

3) RESULTS

Among the four smut fungi genomes compared, *S. scitamineum* and *S. reilianum* were the closest, with an average nucleotide identity of 82.5% among all genes encoding predicted proteins (Table 1). *S. scitamineum* and *S. reilianum* chromosomes displayed syntenic regions, but rearrangements as well as breaking points can be detected. One of such breaking points is a region relevant to the biology of *S. scitamineum* that links the two mating-type *loci*. In *S. scitamineum*, both *loci* are located 59 kbp apart on chromosome 2, characteristic of a bipolar mating system (Figure 1). Chromosome 2 of *S. scitamineum* is homologous to chromosomes 1 and 20 of *S. reilianum*, which also harbor the mating-type *loci b* and *a*, respectively, in a tetrapolar arrangement. Pairwise comparisons of these chromosomes revealed that the breakpoint occurred at the position 825,021 of S. *scitamineum* chromosome 2 (Figure 1). Within the inter-mating-type *loci*, several remnants of DDE_1 and LINE transposons were noted. In addition, 20% of the sequence in between bases 801,345 and 854,308 is composed of repetitive elements, a much larger percentage when compared to the respective 1.24% of the genome.

Table 1. Percentage of identity of mating-type proteins and the average of predicted proteins from whole genome between *S. scitamineum* and others smut fungi.

***S. scitamineum* proteins**	**Percentage identity (BLASTp e-value ≤ 1 x e-14)**		
	S. reilianum	***U. maydis***	***U. hordei***
Average of all predicted proteins	82.5%	75.4 %	72.4%
bE1	68% (bE2)	53% (bE1)	50% (bE1)
bW1	63% (bW2)	43% (bW1)	41% (bW1)
pra1	28% (pra2)	72% (pra1)	69% (pra1)
mfa1.2	Ns* (mfa2.1) Ns* (mfa2.3)	73% (mfa1.2)	Ns* (mfa1.2)
mfa1.3	Ns* (mfa2.1) 66% (mfa2.3)	Ns* (mfa1.2)	Ns* (mfa1.2)

NOTE - The mating-type proteins compared here were those present in the haploid genome sequenced of each smut fungi to allow the comparison with the average identity of proteins from the whole genome.
*Ns - No significant

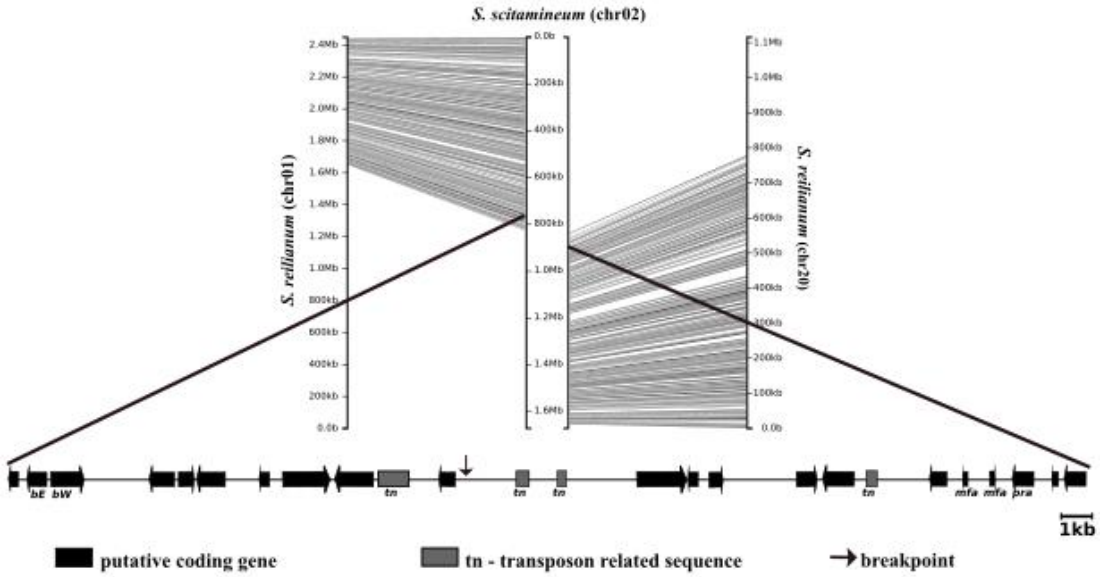

Figure 1. Blocks of synteny between chromosome 2 of *S. scitamineum* and chromosomes 1 and 20 of *S. reilianum* and schematic representation of the linked mating-type *loci* in *S. scitamineum*. The syntenic regions were defined by BLASTn e-value ≤ 1 x 10-5. The region containing the mating-type genes in *S. scitamineum* is expanded. The chromosome breakpoint is identified and indicated by a thin arrow above the sequence. Genes are indicated by black arrows placed according to transcriptional orientation and the transposons related sequences are indicated by grey blocks.

Based on the percentage of sequence identity, phylogenetic analysis and gene organization comparisons (Figures 2, Figure 3 and Table 1), the mating-type proteins of close and distantly related smut species differ substantially. BLASTp analysis shows that predicted proteins from whole genome of *S. scitamineum* have, on average, 82.5, 75.4 and 72.4% of

identity with proteins of *S. reilianum, U. maydis* and *U. hordei*, respectively. In contrast, their mating-type proteins have lower percentages of identity (Table 1).

Phylogenetic analyses were performed including other smut mating-type protein sequences extracted from databases and proteins encoded by the opposite mate pair of *S. scitamineum* obtained by sequencing BAC genomic library (Figure 2). Proteins encoded by *b locus* (*bE1/bE2* and *bW1/bW2*) of *S. scitamineum* (SSC39A and SSC39B) are more close related to each other than to the *bE* and *bW* proteins from related species; whereas proteins encoded by its *a locus* (*pra1/pra2* and *mfa1.2/mfa1.3/mfa2.1/mfa2.3*) cluster preferentially with alleles from others species (Figure 2). The genomic context indicates that the order and orientation of the genes in the *b locus* are conserved (Figure 3). Otherwise for the *a locus*, these settings vary among species and between alleles of the same species, especially by the presence of *rga2* and *lga2* genes in the a2-type alleles that are absent in the a1-type (Figure 3). The *a locus* of *S. scitamineum* and *S. reilianum* also differs from that of *U. maydis* and *U. hordei* by the presence of a second pheromone gene, named *mfa1.3* in *S. scitamineum* due to its similarity to the pheromone *mfa2.3* from *S. reilianum*. It is noteworthy that this extra copy of *mfa* in *S. scitamineum* lacks the characteristic CAAX motif (C, cysteine; A, aliphatic amino acid; X, any amino acid) at its carboxyl terminus. Pheromone response elements (PREs), with a binding motif "ACAAAGGGA" for transcription factor *prf1*, were found close to the mating-type genes in *S. scitamineum* and in the others fungi analyzed, with the exception of *Malassezia* species (Figure 3).

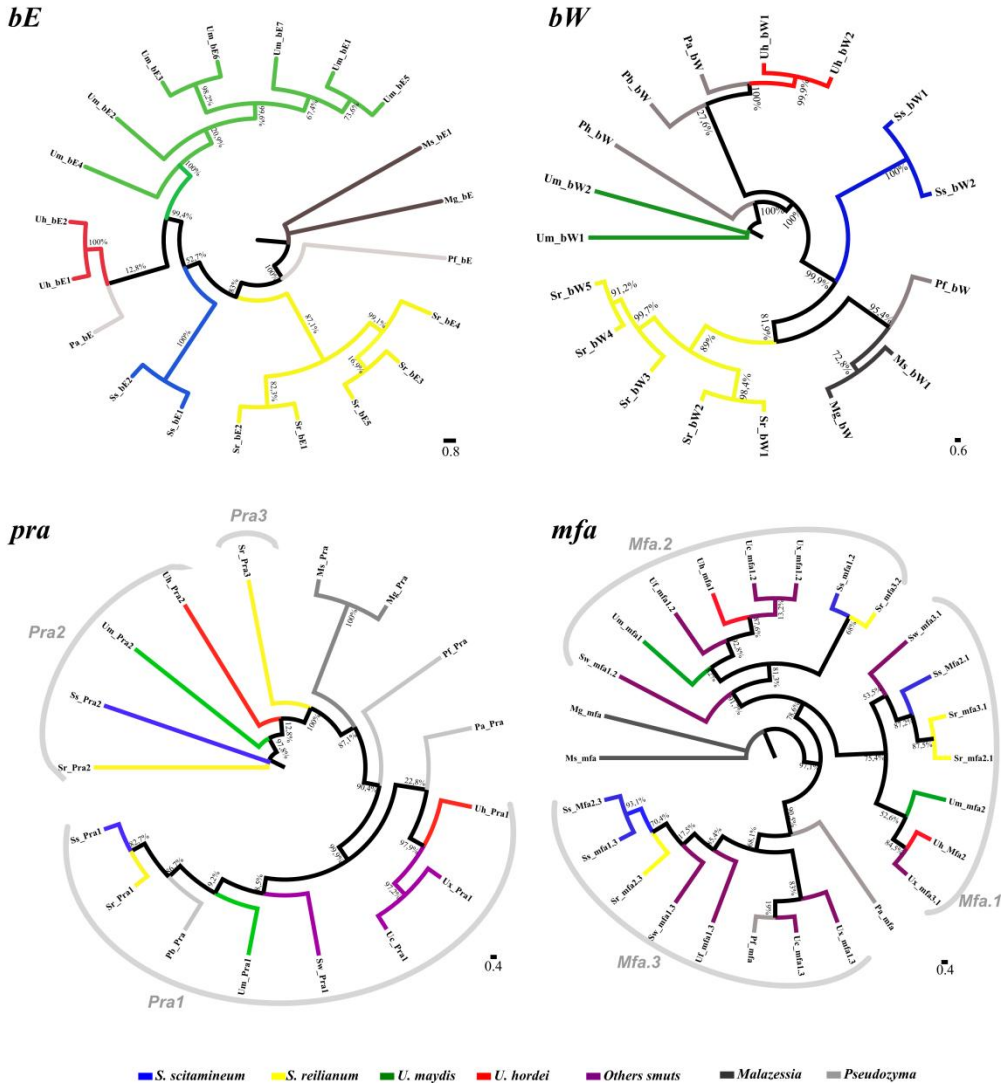

Figure 2. Phylogenetic tree for the mating type proteins from related basidiomycetes species obtained by Maximum Likelihood method implemented in PhyML v3.0. T-COFFEE default alignments, amino acid substitution model provided by ProtTest v.3.2, NNI and aLRT SH-like were used in the analysis. The scale bar indicates the number of amino acid substitutions per site. Acession numbers are: **bE:** Sr_bE1 (CAI59728.1); Sr_bE2 (CAI59732.1); Sr_bE3(CAI59736.1); Sr_bE4 (CAI59740.1); Sr_bE5 (CAI59744.1); Um_bE1 (EAK81226.1); Um_bE3 (P22017.1); Um_bE4 (P22018.1); Um_bE5 (P22019.1); Um_bE6 (P22016.1); Um_bE7 (P22021.1); Uh_bE1 (CAJ42019.1); Uh_bE2 (CAA79216.1); Pa_bE (GAC73813.1); Pf_bE (XP_007878972.1); Ms_bE1 (AGC24186.1); Ms_bE1 (AGC24186.1); Mg_bE (XP_001731616.1). **bW**: Sr_bW1 (CAI59727.1); Sr_bW2 (CAI59731.1); Sr_bW3 (CAI59735.1); Sr_bW4 (CAI59739.1); Sr_bW5 (CAI59743.1); Um_bW1 (EAK81227.1); Um_bW2 (AAA34221.1); Uh_bW1 (CAJ42018.1); Uh_bW2 (CAA79217.1); Pa_bW (AC73812.1); Pf_bW (XP_007878973.1); Ph_bW (GAC99319.1); Pb_bW (EST06506.1); Ms_bW1 (AGC24185.1); Mg_bW (XP_001731615.1). **Pra**: Sr_Pra1 (CAI59749.1); Sr_Pra2 (CAI59755.1); Sr_Pra3 (CAI59763.1); Um_Pra1 (EAK83421.1); Um_Pra2 (AAA99768.1); Uh_Pra1 (CAJ41875); Uh_Pra2 (AAD56044.1); Sw_Pra1 (AEY62480.1); Ux_Pra1 (AEY62504.1); Uc_Pra1 (AEY62485.1); Pa_Pra (GAC74678.1); Pf_Pra (XP_007877412.1); Pb_Pra (EST05906.1); Ms_Pra (AGC13097.1); Mg_Pra (XP_001731696.1); **Mfa**: Sr_mfa2.1 (CAI59758.1); Sr_mfa2.3 (CAI59748.1); Sr_mfa3.1 (CAI59764.1); Sr_mfa3.2 (CAI59762.1); Um_mfa1 (XP_758529.1); Um_mfa2 (P31963.1); Uh_mfa1 (AAC02682.1); Uh_mfa2 (AF184069.1); Uc_mfa1.2 (AEY62483.1); Uc_mfa1.3 (AEY62484.1); Ux_mfa1.2 (AEY62502.1); Ux_mfa1.3 (AEY62503.1); Ux_mfa3.1 (AEY62543.1); Uf_mfa1.2 (AEY62489.1); Uf_mfa1.3 (AEY62490.1); Sw_mfa1.2 (XP_758529.1); Sw_mfa1.3 (AEY62479.1); Sw_mfa3.1 (AEY62524.1); Pa_mfa; Pf_mfa (XP_007877543.1); Ms_mfa (AGC13096.1); Mg_mfa (XP_001731695.1). *S. scitamineum* proteins were obtained from the genome and from the opposite mating-type strain using a BAC library: Ss_bE1 (g971_chr02_Ss); Ss_bE2 (BAC); Ss_bW1 (g970_chr02_Ss); Ss_bW2 (BAC); Ss_Pra1 (g989_chr2_Ss.1); Ss_Pra2 (BAC); Ss_mfa1.2 (g988_chr02_Ss.1); Ss_mfa1.3 (g988_chr02_Ss.2); Ss_mfa2.1 (BAC), Ss_mfa2.3 (BAC).

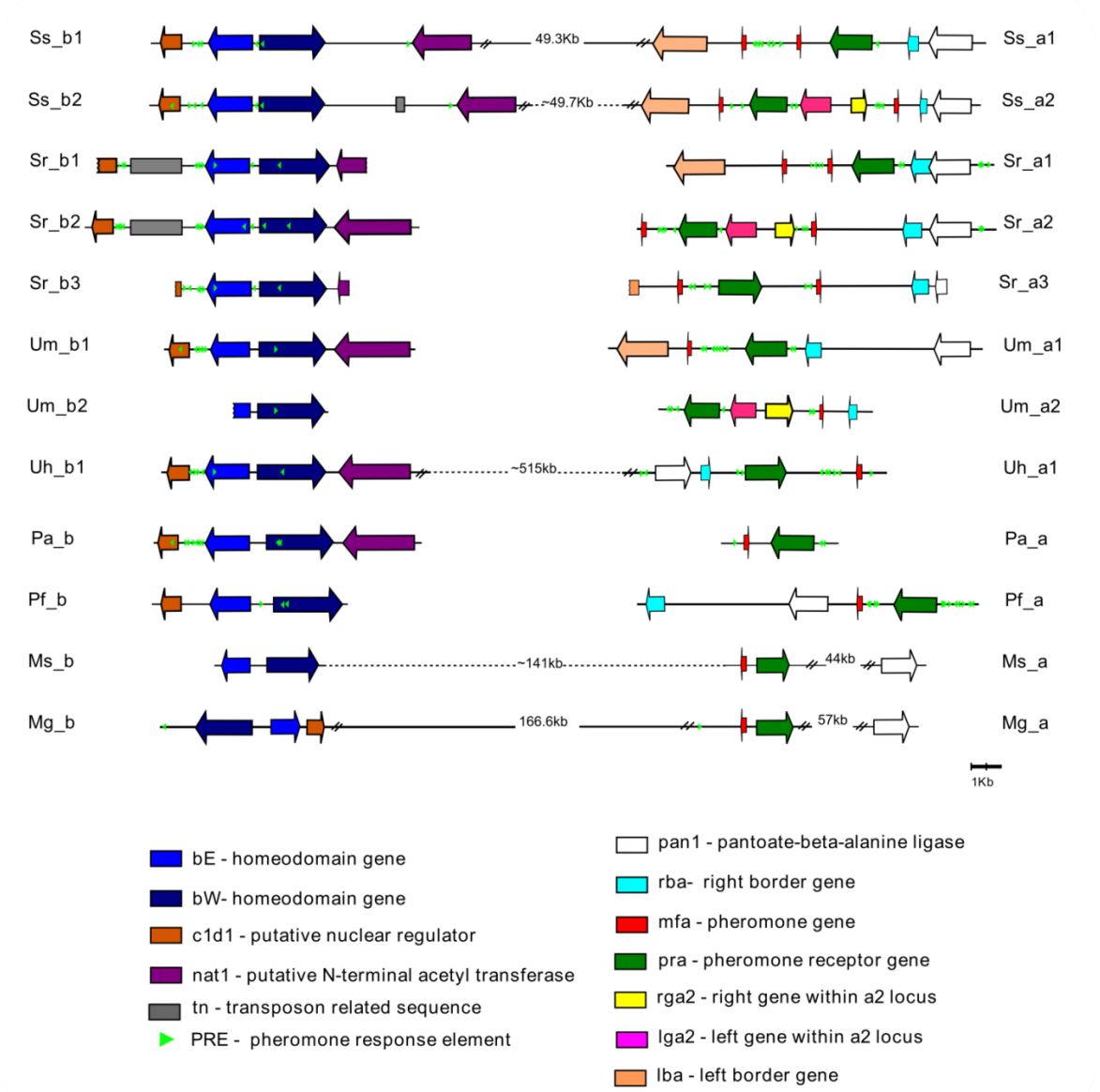

Figure 3. Genomic context of mating-type genes from related fungi species. The species can be identified by the initial letters of their scientific nomenclature (*S. scitamineum*, *S. reilianum*, *U. maydis*, *U. hordei*, *Pseudozyma antarctica*, *Pseudozyma flocculosa*,*Malassezia sympodialis*, *Malassezia globosa*). Genes are indicated by arrows which orientation shows the guidance of transcription. The same colors indicate probable orthologous genes and their functions are listed at the lower position of the figure. The pheromone response elements (consensus sequence ACAAAGGGA) are also represented in the scheme. The distance between linked *loci* is indicated and is not drawn to scale. Accession numbers for **b** ***loci*** were: Sr_b1 (AJ884583); Sr_b2 (FQ311430.1), Sr_b3 (AJ884585), Um_b1 (AACP01000013), Um_b2 (M84182), Uh_b1 (CAGI01000150.1), Pa_b (BAFG01000292.1), Pf_b (AOUS01000449), Ms_b (CANK01000008), Mg_b (AAYY01000003.1). For **a** ***loci***: Sr_a1 (AJ884588.1), Sr_a2 (FQ311442.1), Sr_a3 (AJ884590), Um_a1 (AACP01000083), Um_a2 (UMU37796), Uh_a1 (CAGI01000150.1), Pa_a (BAFG01000390.1), Pf_a (AOUS01000182), Ms_a (CANK01000022), Mg_a (AAYY01000003.1).

4) DISCUSSION

Mating genes are crucial factors for disease establishment since a successful mating reaction is needed to form the infective dikaryotic hyphae. Both processes, mating and pathogenicity, are regulated by two *loci*: *a* and *b* (Bakkeren *et al*., 2008). Although considered to be phylogenetically close based on ITS and LSU rDNA analysis (Stoll *et al*., 2003; Stoll *et al*., 2005), the two genera of smut fungi (*Ustilago/Sporisorium)* show high rates of amino acid substitution per site for mating-type proteins. The trans-specific polymorphisms in pheromone and receptor genes are supposedly preserved since the last common ancestor of the basidiomycetes and ascomycetes and their reciprocal specificity likely have co-evolved (Kellner *et al*., 2011). The pheromone and receptor proteins differ among *Ustilaginaceae* species and seem to be optimized for intraspecific compatibility. However, interspecific sex up to the stage of plasmogamy can be still observed *in vitro*, which could have an evolutionary impact on speciation by hybridization events (Kellner *et al*., 2011).

The genomic context of mating-type genes has been published for several smut fungi (Que *et al*., 2014; Bakkeren *et al*., 2008; Kellner *et al*., 2011; Schirawski *et al*., 2005). The organization of the *b locus* genes is conserved among species in opposition to what has been observed for the *a locus*. As pheromone and receptor genes are separated by a region unique to each mating type, another barrier, beyond the physical distance, suppresses the recombination and avoids the emergence of self-compatible haploids cells (Devier *et al*., 2009).

S. scitamineum and *S. reilianum* differ from *U. maydis* and *U. hordei* by the presence of a second pheromone gene (*mfa)* in that *locus*. However, this extra copy (*mfa 1.3*) in *S. scitamineum* lacks the characteristic sequence acting as a signal for post-translational processing. The post-translational event involves the isoprenylation and carboxymethylation in the cysteine residue to form a secretable lipopeptide pheromone (Caldwell *et al*., 1995). In *U. maydis*, a remnant of a pheromone gene is described for the *a2* allele that cannot produce a functional product (Urban *et al*., 1996). Although *mfa1.3* gene of *S. scitamineum* has full coverage of RNAseq reads, additional experiments are required to determine how many different mating-type alleles are in the fungal population and whether both pheromones are functional.

In *S. reilianum* and *U. maydis*, the mating-type *loci* are not linked and segregates independently (tetrapolar system) (Schirawski *et al*., 2005); in *U. hordei* and *S. scitamineum* the *a* and *b loci* are linked and segregate as one *locus* (bipolar system) (Que *et al*., 2014;

Laurie *et al*., 2012). Here we present the first completely sequenced intergenic region between the two mating *loci* among the smut genotypes and the occurrence of transposable elements at this region. In *U. hordei*, that also has a bipolar mating system, the scaffold that holds the mating-type *loci* has large stretches of long terminal repeats (LTRs) and transposable elements (TEs) dispersed within an intervening region of about 500 kbp (Laurie *et al*., 2012). The TEs may have been responsible for the evolutionary process that resulted in the fusion of mating-type *loci* and the divergence in the inter*loci* segments that have led to a suppression of recombination between the alleles (Bakkeren *et al*., 2008). The tetrapolarity is known only for basidiomycetes and is likely to be an ancestral characteristic since it is observed in at least two of the three major lineages (at the smuts and the mushrooms fungi). The transition to bipolarity must have occurred several times independently during evolution via chromosomal linkage (Bakkeren, Kämper and Schirawski, 2008; Kellner *et al*., 2011). A selective advantage of the bipolarity is the shift from outbreeding to inbreeding lifestyles that could allow the adaptation to specific niches, for example, a novel host.

Transcriptome analysis of *S. scitamineum* during *in vitro* growth and development *in planta* showed that mating-type genes are expressed in both conditions. The two mating-types were grown individually in order to obtain the RNAseq data *in vitro*, which suggests that contact is either not needed to favor gene expression of mating-type genes, or that a very short period of interaction is enough to induce gene expression, since cells were mixed immediately prior centrifugation and RNA extraction (Taniguti *et al*., 2015). The expression of mating-type genes at the *a* and *b loci* are induced by pheromone in *U. maydis*, that leads to amplification of the pheromone signal during the mating and to increased expression of *b* genes prior to fusion (Hartmann *et al*., 1996). A single HMG box transcription factor, PrfI (pheromone response factor I), plays an important role in mediating this interaction between the *a* and *b* pathways. Prf1 protein binds specifically to pheromone response elements (PREs), which occur in clusters in the promoter regions in the vicinity of all genes in both *loci* (Hartmann *et al*., 1996). Although in different numbers and orientations, PREs have been found close to the mating-type genes in *S. scitamineum* and other fungi studied, excepting *Malassezia* species to which so far no sexual cycle has been observed (Saunders *et al*., 2012). The *pfr1* gene is also expressed in all RNAseq experiments analyzed of *S. scitamineum*, suggesting that the same kind of regulation of *a* and *b loci* occurs in this species.

5) CONCLUSION

The complete genome assembly of *S. scitamineum* allowed the resolution of the intergenic region between the two mating-type loci, confirming its bipolar system and revealing the presence of transposable elements as potential drivers of chromosomal rearrangements. The degree of divergence between mating-type alleles among and within species was also assessed. The *a* locus showed an interesting pattern of divergence between alleles, suggesting that interspecific sex could occurs. The presence of a third pheromone allele in *S. scitamineum* remains to be investigated by functional and populational experiments.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Bakkeren G, Kämper J, Schirawski J (2008). Sex in smut fungi: Structure, function and evolution of mating-type complexes. *Fungal Genetics and Biology* **45**:S15–S21.

Billiard S, López-Villavicencio M, Hood ME, Giraud T (2012). Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *Journal of Evolutionary Biology* **25**: 1020–1038.

Caldwell GA, Naider F, Becker JM. (1995). Fungal lipopeptide mating pheromones: a model system for the study of protein prenylation. Microbiological Reviews **59**(3):406–22.

Darriba D, Taboada GL, Doallo R, Posada D (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics **27**(8):1164–1165.

Devier B, Aguileta G, Hood ME, Giraud T (2009). Ancient trans-specific polymorphism at pheromone receptor genes in Basidiomycetes. Genetics **181**: 209–223.

Fraser JA and Heitman J (2003). Fungal mating-type loci. Current Biology **13**: 792–795

Giraud T, Yockteng R, López-Villavicencio M, Refrégier G, and Hood ME (2008). Mating system of the anther smut fungus *Microbotryum violaceum*: selfing under heterothallism. Eukaryotic Cell **7**: 765–775.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology **59**(3):307–321.

Hartmann HA, Kahmann R, Bölker M (1996). The pheromone response factor coordinates filamentous growth and pathogenicity in *Ustilago maydis*. EMBO Journal **15**(7):1632–1641.

Kaltz O and Shykoff JA (1999). Selfing versus outcrossing propensity of the fungal pathogen *Microbotryum violaceum* across *Silene latifolia* host plants. Journal of Evolutionary Biology **12**: 340–349.

Kellner R, Vollmeister E, Feldbrügge M, Begerow D (2011). Interspecific Sex in Grass Smuts and the Genetic Diversity of Their Pheromone-Receptor System. PLoS Genetics **7**(12):e1002436.

Laurie JD, Ali S, Linning R, Mannhaupt G, Wong P, Guldener 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. The Plant Cell **24**:1733–1745.

Notredame C, Higgins DG, Heringa J (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology **302**(1):205–17

Que Y, Xu L, Wu Q, Liu Y, Ling H, Liu Y, *et al.* (2014). Genome sequencing of *Sporisorium scitamineum* provides insights into the pathogenic mechanisms of sugarcane smut. BMC Genomics **15**(1):996.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA and Barrell B. (2000). Artemis: sequence visualization and annotation. Bioinformatics **16**(10): 944-995.

Saunders CW, Scheynius A, Heitman J. (2012). Malassezia Fungi Are Specialized to Live on Skin and Associated with Dandruff, Eczema, and Other Skin Diseases. PLoS Pathogens. **8**(6):e1002701.

Schirawski J, Heinze B, Wagenknecht M, Kahmann R (2005). Mating Type *Loci* of Sporisorium reilianum: Novel Pattern with Three a and Multiple b Specificities. Eukaryotic Cell **4**(8):1317–1327.

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. Canadian Journal of Botany **81**(9):976–984.

Stoll M, Begerow D, Oberwinkler F. Molecular phylogeny of Ustilago, Sporisorium, and related taxa based on combined analyses of rDNA sequences (2005). Mycological Research **109**(3):342–56.

Taniguti LM, Schaker PDC, Benevenuto J *et al*. (2015). Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with sugarcane. PLoS One **10**: e0129318.

Urban M, Kahmann R, Bölker M (1996). The biallelica mating type *locus* of Ustilago maydis: remnants of an additional pheromone gene indicate evolution from a multiallelic ancestor. Molecular and General Genetics **250**(4):414–420.

CHAPTER 4: The genetic basis of host specialization in smut fungi

ABSTRACT

Host specialization is a key evolutionary process for the diversification and emergence of new pathogens. However, the molecular determinants of host range are largely unknown. Smut fungi are biotrophic pathogens that infect mainly Poaceae hosts, including agriculturally important crops, such as cereals, sugarcane, and forage grasses. Despite being phylogenetically close, species of smut fungi have distinct and narrow host ranges. Smut fungi are thought to become specialized following host jumps, as the host and pathogen phylogenies show incongruent topologies. Hence, we aim to identify the genetic basis of host specialization in smut fungi using comparative genomics analyses. We analyzed nine species of smut fungi isolated from eight hosts (crop and non-crop plants): maize, barley, sugarcane, wheat, oats, *Zizania latifolia* (Manchurian wild rice), *Echinocloa colona* (a wild grass), and *Persicaria sp*. (a wild dicot plant). We assembled two new genomes: *Ustilago hordei* isolated from oats and *Ustilago tritici* isolated from wheat. The smut genomes were of small sizes, ranging from 18.38 to 24.64 Mb, had low repetitive element content (2.34 – 25.12%), and low number of duplicated genes (6,5 – 12,6% of the genes had a paralog). *Ustilago hordei* species seem to be experiencing genome expansion due to the proliferation of transposable elements with varying amounts even between the two strains. The smut species shared 65.46% of all orthologous groups as defined by OrthoMCL analyses. The CAZyme, protease and lipase content were similar among the species. Most of the species-specific genes had no functional domain assigned, including genes encoding candidate effectors. All one-to-one orthologous genes were screened for positive selection. Using site-specific models, we found significant evidence of positive selection at individual sites in 31 genes. Using branch-site models, we found evidence for episodic selection in 41 to 286 genes, depending on the analyzed lineage. The genetic basis of host specialization in smut fungi is complex and seems to involve a range of evolutionary processes, including gene gain/loss and episodic selection events. Species-specific effectors and positively selected genes will be good candidates for further characterization in regards to their role in host adaptation.

Keywords: Ustilaginaceae; Host jump; Comparative genomics; Effectors; Orphan genes; Positive selection

1) INTRODUCTION

Host specialization is a common strategy among plant pathogens. A general prediction is that specialists outperform generalists on a given host species at the expense of the ability to successfully infect alternative hosts (the ‘jack of all trades – master of none’ principle) (Barrett and Heil, 2012). Specialist pathogens are favored in ecological contexts of restricted host species diversity, interspecific competition, and due to genetic trade-offs in adaptation to different hosts (Johnson et al., 2009; Barrett et al., 2009). Moreover, the co-evolutionary process itself is conducive to ever-increasing host specialization. The strong host selective pressure is likely to result in more specialized pathogen lineages over time and

phylogenetically restricted host ranges (Gilbert and Webb, 2007; Johnson et al., 2009; Antonovics et al., 2012).

The intimate interaction between plants and specialist pathogens suggests that co-speciation should be common. However, phylogenetic congruence of host and pathogens phylogenies is rare. Host shifts/jumps rather than co-speciation are the main mode of pathogen speciation and a major route for disease emergence (Giraud et al., 2008; Giraud et al., 2010; de Viene et al., 2013; Choi and Thines, 2015). This raises intriguing questions such as how do specialized pathogens shift and specialize on a novel host, and which are the genetic determinants of host specificity. Different hosts have different defense mechanisms, biochemical composition, and associated microbiota to which pathogens must adapt to in order to be able to infect, colonize, feed and reproduce (Barret and Heil, 2012; Walker et al., 2013; Haueisen and Stukenbrock, 2016). Hence, specialization to any specific host likely requires a different set of adaptations.

Many pathogens show extraordinary genome plasticity and this most likely generates the genetic variability that enables pathogen genotypes quickly respond to selection pressures imposed by a new host (Plissonneau et al., 2017). Analyses of host adaptation processes through comparative genomic studies showed that gene gain/loss, gene family expansion/contraction, and adaptive mutations were the most likely mechanisms across different pathosystems (Raffaele et al., 2010; Ma et al., 2010; Burmester et al., 2011; Zhong et al., 2016; Yoshida et al., 2016; Grandaubert et al., 2015; Poppe et al., 2015; Baltrus et al., 2012; Kirzinger and Stavrinides, 2012). Given that different genetic basis are involved in host specialization and many of the genomic changes may be unrelated to the process of host adaptation, it is crucial to study each pathogen lineage and consider as many of the most closely related species as possible.

Smut fungi are suitable plant pathogens to investigate host specialization mechanisms. Despite the growing interest in smut diseases as a threat to agriculture, edible delicacies, and biotechnological applications (Feldbrügge et al., 2013; Toh and Perlin, 2016), the genetic basis of host specialization in smut fungi remains largely unknown. True smuts belong to the phylum Basidiomycota, class Ustilaginomycetes, and comprise more than 1,650 species. Smut species infect hosts from many angiosperm clades. However, most of smut species are highly specialized on a single or a few host species, affecting mainly members of Poaceae family (Begerow et al., 2004). Smut fungi are thought to become specialized following host jumps, as the host and pathogen phylogenies show incongruent topologies (Begerow et al., 2004). The estimated divergence dates of four smut pathogens from agronomically important

crops support the hypothesis that the host specialization evolved after the speciation of the host, but before the domestication of the host (Munkacsi et al., 2007).

Smut diseases are characterized by the production of a sooty dark brown mass of teliospores (Bakkeren et al., 2008; Morrow and Fraser, 2009). The life cycle comprises three genetically and morphologically distinct phases: diploid teliospores, haploid yeast like-cells and dikaryotic infective hypha (Piepenbring, 2009). Despite of their similarities, the local and mode of plant infection and symptom development also vary among species. For example, *Ustilago maydis*, the causal agent of common smut of maize and teosinte, infects all aerial parts of the host plant (stems, leaves, tassels and ears) and locally induces tumor formation (Bölker, 2001; Matei and Doehlemann, 2016); while most of smut species become systemic and the symptoms occurs only in floral tissues (Piepenbring, 2009). The local of infection also varies among species, with some penetrating through the ovary, coleoptile, leaves, roots or young buds. A common secondary symptom of many smut diseases is the hypertrophy of specific host organs, forming tumor-like galls (Piepenbring, 2009). The loss of apical dominance is also described for some species, inducing the formation of multiple female inflorescences in *Sporisorium reilianum* infecting maize (Ghareeb et al., 2015) and tillering in *S. reilianum* infecting sorghum (Matheussen et al., 1991) and *S. scitamineum* infecting sugarcane (Sundar et al., 2012).

In order to investigate the genetic basis of host specialization, we performed a broad comparative genomic study of smut fungi. We used the genome sequences of seven smut species from Ustilaginaceae family available in databases to date. We additionally sequenced the genomes of two species isolated from agronomically important crops (wheat and oats) to increase the host range survey. A total of nine smut pathogens isolated from eight distinct hosts were compared, including three species infecting wild hosts (*Zizania latifolia*, *Echinochloa colonum*, and *Persicaria sp*.) and six isolates from domesticated crops (maize, barley, oats, wheat, sugarcane). We looked for differences in genomic features and gene content between the species, paying special attention to candidate cell-wall degrading enzymes and effectors due to their known role in plant-pathogen interactions. We also focused on species-specific genes and genes under positive selection as potential drivers of host specialization. Our results provide good candidate genes for functional characterization.

2) MATERIAL AND METHODS

2.1) Strains, DNA extraction and sequencing

For genome sequencing, we selected the *Ustilago hordei* (strain Uhor01) isolated from an oats field in Southern Brazil and the *Ustilago tritici* (syn. *Tilletia tritici*) from CBS database (strain Utri01). Yeast-like cells were obtained from *U. hordei* teliospores according to Albert and Schenck (1996). For genomic DNA extraction, single colonies from both species were grown in YM liquid medium (0.3% yeast extract, 0.3% malt extract, 0.5% soybean peptone, 1% D-glucose), overnight, at 25 °C, in an orbital shaker. Genomic DNA was extracted using the Genomic-tip 20G kit (Qiagen, Inc.), according to manufacturer's instructions for yeasts. A total of 10 μg of DNA of each sample was sent to the GCB facility at Duke University (USA), where a single large insert library (15kb-20kb) was constructed and sequenced in one SMRT cell (P5-C3 chemistry) using the PacBio RS II (Pacific Biosciences, Inc.) sequencing platform. DNA from the same extraction was also used for Illumina paired-end library construction and sequencing using HiSeq2500 platform with 2x100 cycles at Center of Functional Genomics (ESALQ/USP, Brazil). For *U. hordei*, we obtained about 10.4 Gb and 4.7 Gb of IIumina and Pacbio data, respectively. For *U. tritici*, we obtained about 1.6 Gb and 0.5 Gb of Illumina and Pacbio data, respectively.

The genome and annotation files of *Ustilago maydis*, *U. hordei*, *Sporisorium reilianum* were retrieved from MIPS (http://mips.helmholtz-muenchen.de/). The sequences of *Ustilago esculenta*, *Ustilago trichophora, Sporisorium scitamineum* were retrieved from NCBI (http://www.ncbi.nlm.nih.gov), and sequences of *Melanopsichum pennsylvanicum* from Senckenberg Repository (http://dx.doi.org/10.12761/SGN.2014.3). Information about the smut and outgroup species used in the present study are listed in Table 1.

Table1. Ustilaginomycotina species and genomes used in this work.

	Abbr.	Species	Strain	Host Source	Accession Number	Project Number	Reference
SMUTS	*UhoO*	*Ustilago hordei*	Uhor01	*Avena sativa* (oats)	-	-	This work
	Utri	*Ustilago tritici syn. Tilletia tritici*	Utri01	*Triticum spp*. (wheat)	-	-	This work
	Umay	*Ustilago maydis*	521	*Zea mays* (maize)	GCA_000328475.2	PRJNA1446	Kamper *et al.* 2006
	UhoB	*Ustilago hordei*	Uh4857-4	*Hordeum vulgare* (sorghum)	GCA_000286035.1	PRJEA79049	Laurie *et al.* 2012
	Uesc	*Ustilago esculenta*	MMT	*Zizania latifólia* (wild rice-relative)	GCA_000819925.1	PRJNA263330	Ye *et al.*, 2014 (unpublished)
	Utcp	*Ustilago trichophora*	RK089	*Echinochloa colona* (wild grass)	GCA_001654535.1	PRJNA316802	Zambanini *et al.*, 2016
	Srei	*Sporisorium reilianum*	SRZ2	*Zea mays* (maize)	GCA_000230245.1	PRJNA64587	Schirawski *et al.* 2010
	Ssci	*Sporisorium scitamineum*	SSC39B	*Saccharum spp*. (sugarcane)	GCA_001010845.1	PRJNA275631	Taniguti *et al.* 2015
	Mpen	*Melanopsichium pennsylvanicum*	Mp4	*Persicaria sp*. (wild dicot plant)	Not avaiable	PRJEB4565	Sharma *et al.* 2014
OUTGROUPS	*Mglo*	*Malassezia globosa*	CBS7966	Human	GCA_000181695.1	PRJNA18719	Xu *et al.* 2007
	Msym	*Malassezia sympodialis*	ATCC42132	Human	GCA_000349305.2	PRJEB417	Gioti *et al.* 2013
	Pant	*Pseudozyma antarctica syn. Moesziomyces antarcticus*	JCM10317	Non-pathogenic	GCA_000747765.1	PRJNA302316	Morita *et al.* 2014
	Paph	*Pseudozyma aphidis syn. Moesziomyces aphidis*	DSM70725	Non-pathogenic	GCA_000517465.1	PRJNA215967	Lorenz *et al.* 2014
	Pbra	*Pseudozyma brasiliensis syn. Kalmanozyma brasiliensis*	GHG001	Non-pathogenic	GCA_000497045.1	PRJNA217085	Oliveira *et al.* 2013
	Pflo	*Pseudozyma flocculosa syn. Anthracocystis flocculosa*	PF-1	Non-pathogenic	GCA_000417875.1	PRJNA185206	Lefebvre *et al.* 2013
	Phub	*Pseudozyma hubeiensis*	SY62	Non-pathogenic	GCA_000403515.1	PRJDB993	Konishi *et al.* 2013

2.2) Genome Assembly

We evaluated multiple approaches for the *de novo* assembly of the *U. hordei* and *U. tritici* genomes. A hybrid assembly using AHA from the SMRT Analysis 2.3.0 (Chin *et al.* 2013) produced the best assembly metrics for both species. To further improve the assembly, PBJelly from the PBSuite v15.8.24 (English *et al.* 2014) was used to fill intra-scaffold gaps in the AHA hybrid assembly through the alignment of long PacBio reads. Pilon v1.18 (Walker *et al.* 2014) was also used to align Illumina short reads to the draft assembly in order to correct single base errors, minor mis-assemblies and to fill gaps.

2.3) Gene Prediction and Annotation

Genes in the genomes of *U. hordei*, *U. tritici*, *U. esculenta*, and *U. trichophora* were predicted using Augustus v.2.5.5 (Stanke and Morgenstern 2005). Protein sequences of *U.maydis*, *U. hordei* and *S. scitamineum* were used as extrinsic sources of gene structure evidence to improve sensitivity of gene predictions. For this, exonerate v.2.2.0 (Slater and Birney 2005) was used to generate hints from protein sequence alignments (protein2genome option).

All predicted proteomes were annotated using InterProScan v.5.19 (Jones *et al.* 2014). Pfam protein families, InterPro domains, gene ontology (GO) classification, and metabolic pathways were recovered. The predicted secretome was defined by the presence of a signal peptide and absence of any transmembrane domain, using Phobius v.1.01 (Käll *et al.* 2004) and SignalP v4.1 (Bendtsen *et al.* 2004). EffectorP was used to predict the effector repertoire from the predicted secretome based on machine learning (Sperschneider *et al.* 2016). Characterized effectors in smut species were screened for orthologs and tblastn was used to search for homologous regions in smut genomes.

The proteomes were also screened for CAZymes (carbohydrate active enzymes) (Lombard *et al.* 2013) using Hmmscan from the HMMER v3.1b2 package (http://hmmer.org/) and the dbCAN HMM profile database (Yin *et al.* 2012). The hmmscan-parser script provided by dbCAN was used to select significant matches. Searches for lipases were also performed with Hmmscan using the "Lipase Engineering Database" (Fischer and Pleiss 2003). Putative peptidases were identified by using batch BLAST at the MEROPS server (Rawlings *et al.* 2012). The secondary metabolite biosynthesis clusters were predicted by AntiSMASH web version 4.0.0 (Medema et al. 2011). Gene Ontology (GO) terms distribution of gene sets were performed with Blast2GO v2.7.2 (Conesa et al., 2008).

2.4) Repeats and Transposable Elements

De novo and homology-based identification of repeats were performed using the RepeatModeler pipeline. A combined repeat library was constructed concatenating the RepBase library (release of August 2015) with the *de novo* repeat family predictions. The combined repeat library was used as input for RepeatMasker (http://www.repeatmasker.org/).

2.5) Orthologous groups

Orthologous and paralogous groups among the nine 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 was parsed to separate core and unique clusters, singletons, single-copy, and one-to-one orthologous genes. Orphan genes included singletons (genes not assigned to any OrthoMCL group) and unique clusters (cluster of paralogs unique to one species). For the phylogenetic tree reconstruction, OrthoMCL was also performed including the genome of additional *Ustilaginomycotina* fungi: *Malassezia globosa*, *Malassezia sympodialis*, *Pseudozyma antarctica*, *Pseudozyma aphidis*, *Pseudozyma brasiliensis*, *Pseudozyma flocculosa*, *Pseudozyma hubeiensis* (for references see Table1).

2.6) Phylogenetic tree

A total of 1,776 one-to-one orthologous proteins from 16 genomes were aligned using MUSCLE v.3.6 (Edgar 2004). Gblocks v.0.91b (Castresana 2000) was used to remove all gaps (-b5=n) and blocks with length smaller than 5 (-b4=5) in each alignment. After Gblocks filtering, protein alignments smaller than 100 amino acids were excluded. 1,637 protein alignments were retained and concatenated for a total 624,996 amino acid positions. The best-fit amino acid substitution model for the data was obtained using ProtTest v.3.4.2 (Darriba *et al.* 2011). The model of LG+I+G+F was selected based on the likelihood and Bayesian criteria. A maximum likelihood phylogenetic tree was constructed using RAxML v.8.2.8 (Stamatakis 2014) with 100 rapid bootstrap partitions. The RAxML super matrix tree was scored by each individual gene tree using ASTRAL v.4.10.8 (Sayyari and Mirarab 2016).

2.7) Positive selection

A total of 4,374 protein-coding sequences with one-to-one orthologs among the nine smut species were aligned with the codon-aware aligner pal2nal v.14 (Suyama *et al.* 2006).

Mismatches with the protein sequence and gaps were removed from the final alignment. After filtering, 4,195 sequence alignments were retained. The protein sequences were used to build a smut phylogenetic tree using the methods described above. The ETE3-evol tool (Huerta-Cepas *et al*. 2016) was used to automate codeml analysis (Yang 2007). Sites and branch-sites models at each lineage were applied. For *U. hordei* lineages, we also consider the species branch for positive selection analysis.

3) RESULTS

3.1) Genome sequence of *U. hordei* and *U. tritici*

We assembled the genomes of two smut species infecting agronomically important crops (*U. hordei* from oats and *U. tritici* from wheat) using a combination of Illumina and Pacbio reads. The *de novo* assemblies resulted in a genome size of 18.64 Mb assembled in 80 contigs (≥ 500 bp) for *U. tritici* and 24.64 Mb assembled in 2,196 contigs (≥ 500 bp) for *U. hordei*. The assembly of *U. tritici* had a higher degree of contiguity (N_{50} 610 kb) than *U. hordei* (N_{50} 40 kb) (Figure 1). A total of 7,892 and 6,776 protein-coding genes were predicted in the *U. hordei* and *U. tritici* genomes, respectively. For both species, the two mating-type loci (*a* and *b*) were located at different scaffolds.

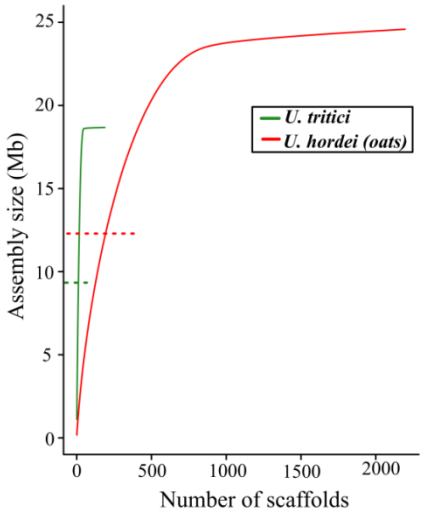

Figure 1. Accumulated length distribution of scaffolds in the newly assembled genomes. Dashed lines represent 50% of the assembled genomes.

3.2) Phylogenomics

To reconstruct the phylogeny of smut fungi we included six additional species from the subphylum Ustilagomycotina. One-to-one protein orthologues were concatenated and used to build a super-matrix tree. The super-matrix tree had a quartet support of 64.80% (i.e. 64.80% of all quartet trees induced from gene trees were present in the super-matrix tree).

The phylogenomic tree showed that the genus *Ustilago* was not monophyletic, clustering along with members of the *Melanopsichum* and *Pseudozyma* genera (Figure 2). The dicot-infecting species, *M. pennsylvanicum*, was closely related to the monocot-infecting pathogens in the phylogenetic tree. The phylogeny of the smut fungi also did not separate pathogens according to the wild or domesticated status of their hosts. *U. hordei* was placed as the most evolutionarily distant species among the analyzed smut fungi.

3.3) Comparative Genomics of Smut Fungi

U. hordei species showed the largest genomes among the smut fungi, which ranged from 18.38 Mb in *S. reilianum* to 24.64 Mb in *U. hordei* (oats) (Table 2, Figure 2). The larger genome size in *U. hordei* species was also accompanied by an increase in the repeat elements content. In particular, the transposable elements content ranged from 0.61% in the genome of *U. tritici* to 23.93% in *U. hordei (oats)*. The predicted gene repertoire varied from 6,280 genes in *M. pennsylvanicum* to 7,892 in *U. hordei* (oats). *M. pennsylvanicum* also had the smallest number of secreted protein (291) and predicted effectors (55), while *S. reilianum* had the largest secretome (443) and effector content (127). The total number and the diversity of sub-categories of CAZyme, protease, and lipase domains were similar among smut species. *U. hordei* species were an exception, because there was an expansion of the peptidase family A11A (*Copia* transposon peptidase) compared to the other species (Figure 3). Around ten secondary metabolite biosynthesis clusters were identified in all smut genomes (data not shown). All species have at least one cluster encoding for putative terpene synthase (TS), non-ribosomal peptide synthase (NRPS), and type 1 polyketide synthases (t1PKS). Only *U. trichophora* presented a hybrid cluster of NRPS-Indole-t1PKS.

Table 2. Genomic statistics of smut fungi.

Genomic Statistics	*UhoO*	*Utri*	*Umay*	*UhoB*	*Uesc*	*Utcp*	*Srei*	*Ssci*	*Mpen*
Assembly									
Total assembly size (Mb)	24,64	18,64	19,64	21,15	20,19	20.68	18,38	19,95	19,23
Average base coverage	487X	278X	10X	25X	139X	na	29X	500X	339X
Number of contigs (>=500bp)	2196	80	27	713	298	215	45	26	435
N50 (bp)	39,442	610,801	884,984	307,727	403,507	179,640	772,363	875,830	121,670
Largest contig (bp)	171,399	1,118,949	2,476,501	542,606	1,882,320	637,988	2,448,206	2,009,762	690,500
GC-content (%)	51.60	57,07	54.03	52.16	54.42	53.06	59.87	55.16	50.90
Coding sequences									
Number of genes	7892	6776	6784	7111	6773	6499	6776	6677	6280
Single-copy genes	6897	6335	6175	6500	6057	5925	6159	6080	5791
In-paralogs	995	441	609	611	716	574	617	597	489
Co-orthologs groups	6319	6055	6111	6351	5808	5890	6214	6007	5591
Genes into the groups	7026	6297	6454	6717	6287	6239	6492	6382	5853
Unique groups (# genes)	82 (321)	8 (21)	12 (37)	21 (53)	32 (121)	17 (39)	8 (18)	19 (61)	7 (20)
Singletons	866	479	330	394	486	260	183	295	427
Total of orphan genes	1.187	500	367	447	607	299	201	356	447
Repeat sequences (%)									
Simple/tandem repeats	1.59	1.75	1.68	1.60	2.12	2.06	2.04	1.76	1.59
Interspersed repeats/TEs	23.93	0.61	2.47	11.68	7.54	2.15	0.66	4.10	2.30
Total of bases masked	25.12	2.34	4.13	13.21	9.65	4.18	2.68	5.85	3.88
Secretome									
Predicted secretome	343	397	441	373	314	394	443	371	291
Predicted effectors (EffectorP)	85	97	124	104	70	96	127	85	55
Predicted effectors (size/Cys)	50	42	58	63	29	42	47	33	20

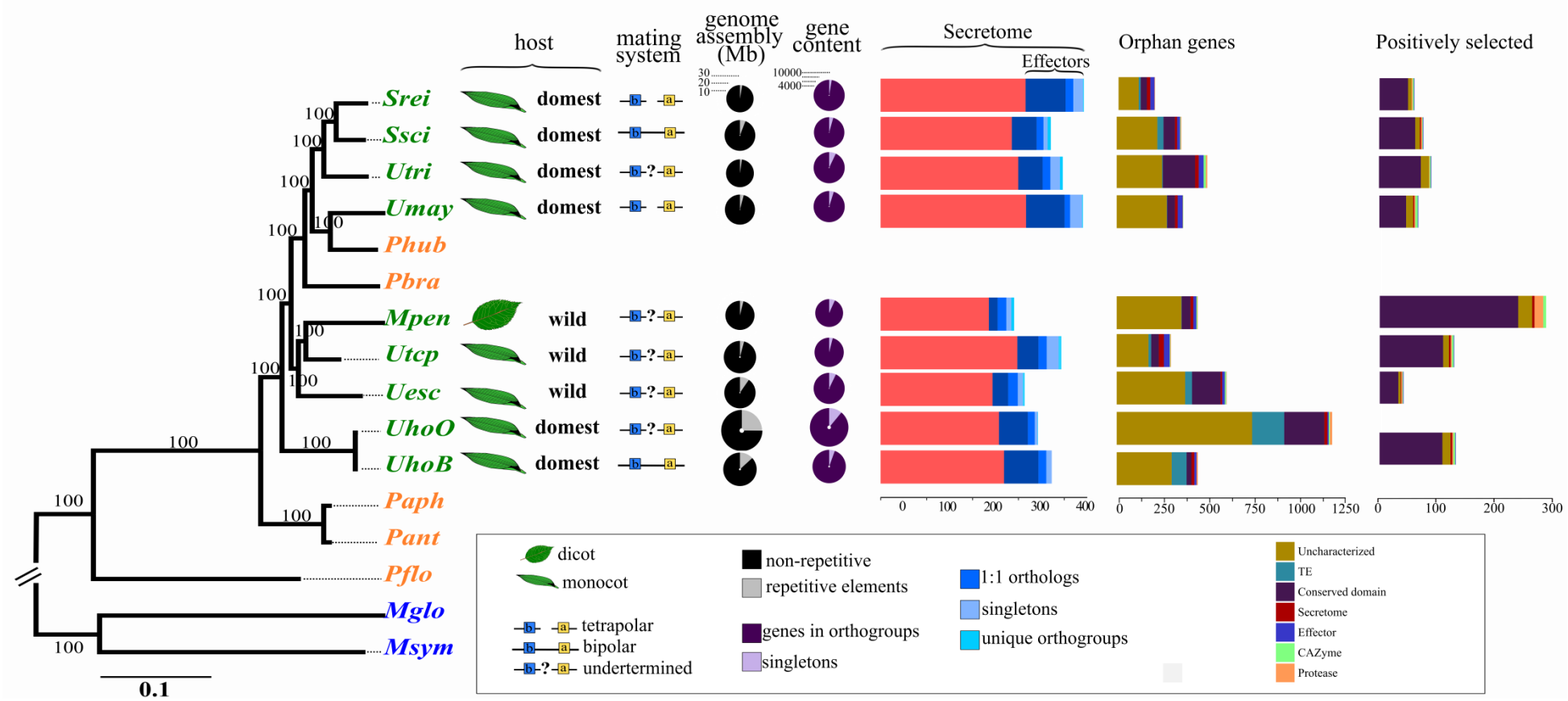

Figure 2. Maximum likelihood phylogenomic tree of *Ustilagomycotina* species based on 1,637 one-to-one orthologous genes, host information, and genomic features of smut fungi. The scientific names were abbreviated according to Table 1. The colors assigned to each species distinguish plant pathogens (green), human pathogens (blue) and non-pathogenic (orange) species. Double bars at the tree root indicate out of scale.

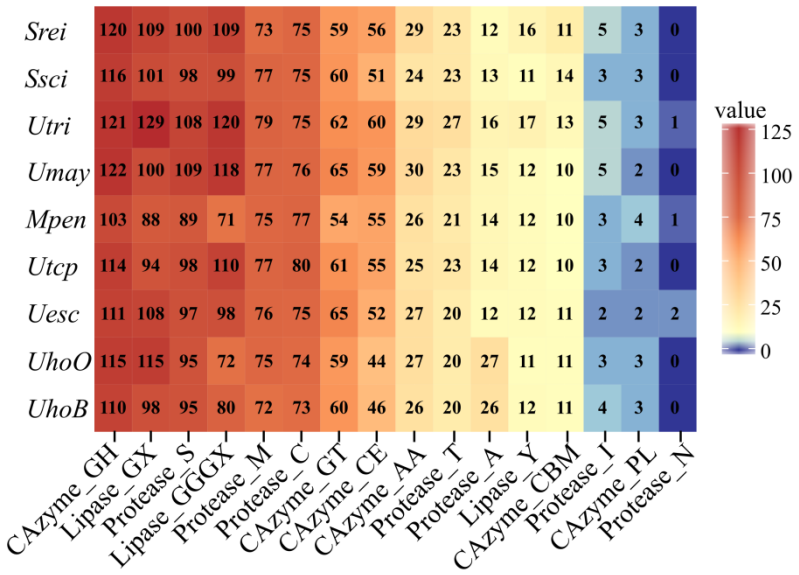

Figure 3. Heatmap of CAZymes, proteases and lipases classes. The numbers of enzyme categories in each genome are shown. Classes and modules of CAZymes include: GHs (Glycoside Hydrolases), CEs (Carbohydrate Esterases), CBMs (Carbohydrate-Binding Modules), GTs (GlycosylTransferases), PLs (Polysaccharide Lyases), AAs (Auxiliary Activities). Proteases are classified by the catalytic type of the proteolytic enzymes: Aspartic (A), Cysteine (C), Metallo (M), Asparagine (N), Serine (S), Threonine (T), and also inhibitors of peptidases (I). Lipases are classified into three classes on the basis of the oxyanion hole: GX, GGGX, and Y.

The predicted proteome from the nine smut species were compared and clustered in 7,187 orthologous clusters (orthogroups) using OrthoMCL. Out of those, 4,706 were shared among all species, wherein 4,374 were one-to-one orthologs. *S. reilianum* had the smallest set of orphan genes (201) and *U. hordei* (oats) the largest (1,187) (Figure 2). Most of the species-specific genes were uncharacterized (without any interpro domain). Transposase enzymes and reverse transcriptases were also found among orphan genes, mainly in *U. hordei* gene sets. Some predicted effectors (ranging from 7 in *U. hordei* (oats) to 32 in *U. trichophora*) were also species-specific. Among the orphan genes with a predicted function, we highlight enzymes acting on primary and secondary metabolic pathways, encompassing both biosynthesis and degradation processes; genes associated with response to stimulus, regulation of gene expression, cell cycle control, sexual reproduction, development, and transport (Figure 4A).

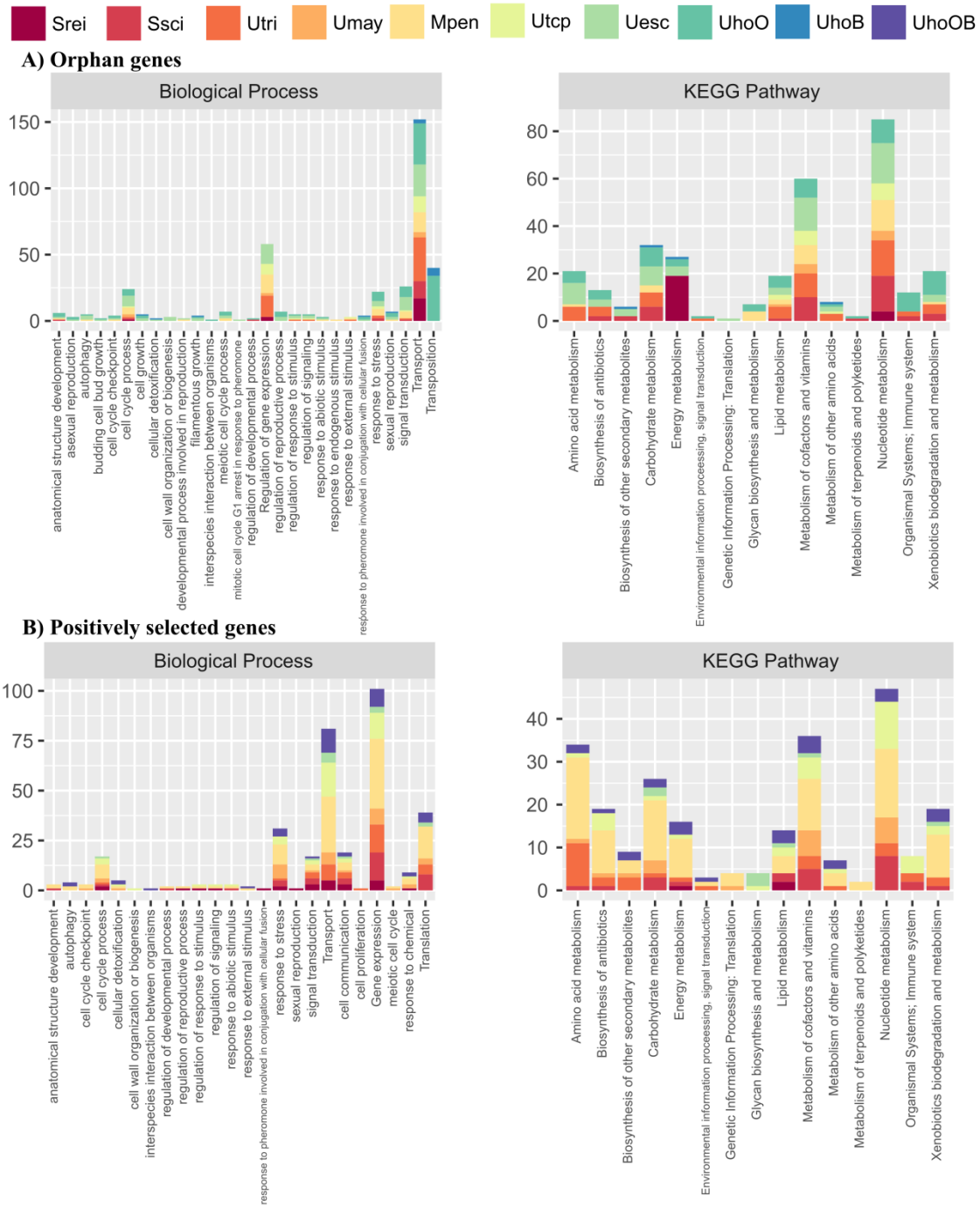

Figure 4. Selected GO terms (Biological Process) and KEGG enzymes pathways for sets of orphan (A) and positively selected (B) genes. The bars are colored per smut pathogen. For the positive selection analyses, the branch for both *U. hordei* species was considered and named as UhoOB. The size of each colored bars represent the number of sequences attributed to each category and it is not cumulative in the overall size of the bar.

Effectors that were well-characterized in *U. maydis*, *S. reilianum* and *U. hordei* (barley) were searched for orthologs defined by OrthoMCL (Figure 5). Some effectors were present in all genomes: Cmu1 (*Chorismato mutase 1*), Stp1 (*stop after penetration 1*), ApB73

(*Apathogenic in B73*), and members of *eff1* family. The leaf-specific expressed effectors candidates, *um06223* and *um12217*, were present only in *U. maydis*. The effector Sad1 (*Supressor of apical dominance 1*) was specific of *S. reilianum* genome. Effectors that were well-characterized in *U. maydis*, *S. reilianum* and *U. hordei* (barley) were searched for orthologs as defined by OrthoMCL. Some effectors were present in all genomes: Cmu1 (*Chorismato mutase 1*), Stp1 (*stop after penetration 1*), ApB73 (*Apathogenic in B73*), and members of *eff1* family; while others showed different patterns of presence/absence (Figure 5). The leaf-specific expressed effectors candidates, *um06223* and *um12217*, were present only in *U. maydis*. The effector Sad1 (*Supressor of apical dominance 1*) was specific of *S. reilianum* genome. Tblastn searches identified similar genomic regions for some "absent" effectors (Figure 5), providing evidences of pseudogeneization events, such as for Pep1 in *U. esculenta*, or automatic gene prediction errors, such as for Sad1 in *S. scitamineum* (data not shown).

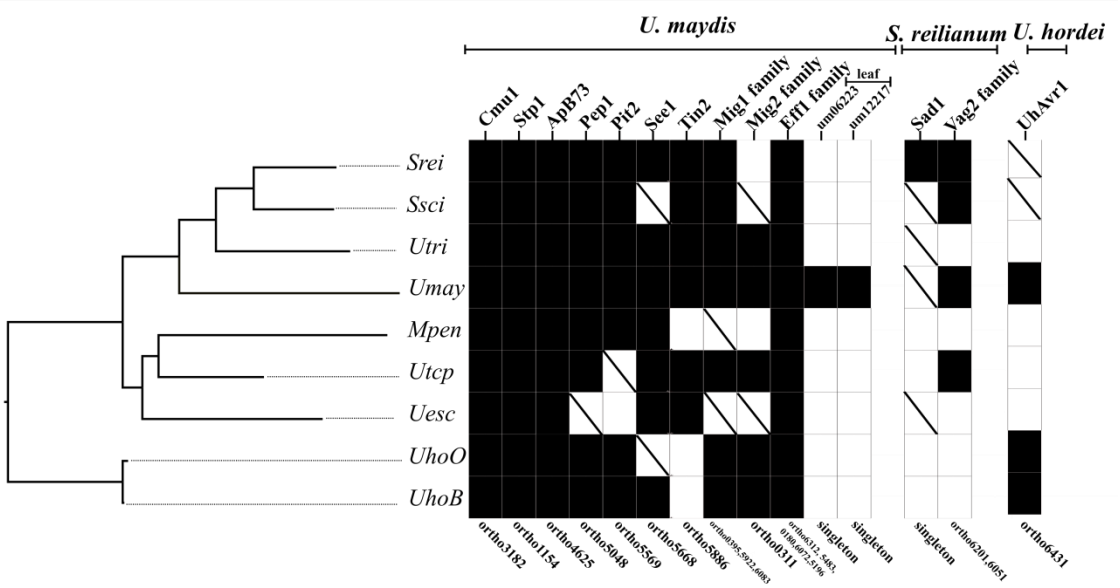

Figure 5. Presence (black squares) and absence (blank squares) of genes or gene families encoding known effectors in *U. maydis*, *S. reilianum* and *U. hordei* based on OrthoMCL groups. The OrthoMCL groups are indicated in the lower part of the figure. Crossed blank squares indicate that a similar genic region is present in the genome by tblastn search. The species are ordered according to their phylogenetic relationship.

3.4) Positively selected genes

The average dN/dS ratio (ω) over the whole coding region showed evidence of purifying selection for all one-to-one orthologous genes. The maximum ω value was 0.418 (Figure 6).

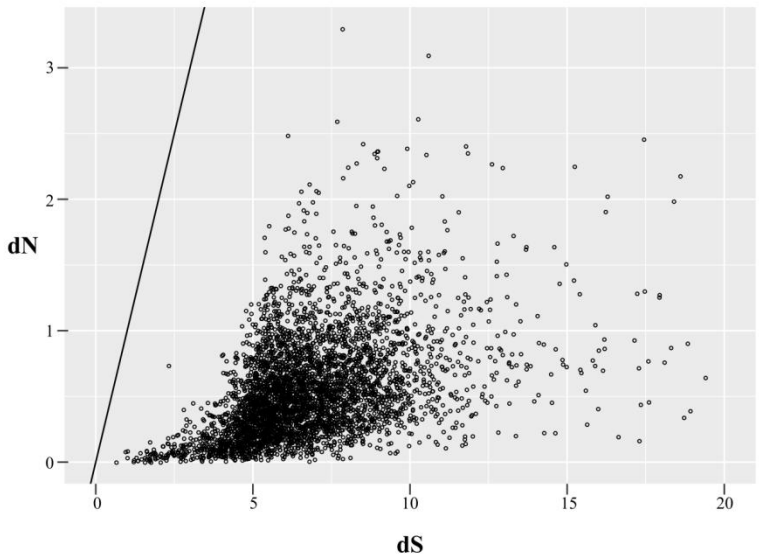

Figure 6. Plot of dN and dS ratios for 4,195 one-to-one orthologous genes. Values were calculated considering the whole coding sequence. The line indicates dN/dS ratio equals 1, representing the threshold between positive (above the line) and purifying selection (below the line).

Using a more sensitive approach based on site-specific models, we found significant evidence of positive selection at individual sites in 31 genes (Supplementary Table 1). These genes were confirmed by both M2/M1 and M8/M7 model comparisons. Out of the 31 genes, three genes were uncharacterized. Interesting domains included those associated with the regulation of transcription, such as the *bE* mating-type-specific homeodomain and synthesis of lipids, such as diacylglycerol acyltransferase domain.

Branch-site models were also used to identify sites evolving under episodic selection. For *U. hordei* lineages analyzed individually, only nine genes in each lineage were detected under selection. Therefore, we consider the *U. hordei* species branch for comparison. Positively selected sites exclusively at one species varied from 41 in *U. esculenta* to 286 in *M. pennsylvanicum*. Members of CAZYmes and proteases categories were found under episodic selection. Among those, *M. pennsylvanicum* showed the highest number, with 5 CAZYmes from esterases families and 15 protease genes from distinct categories under positive selection (Figure 2). General functional information about positively selected genes is also present in Figure 4B.

4) DISCUSSION

4.1) Complex evolution of smut fungi: Taxonomic and gene tree discordances

The term "smut fungi" refers to a non-monophyletic group of plant pathogens that produce a characteristic dark spore mass (teliospores). Smut fungi belonging to Ustilaginomycetes class form a monophyletic group known as "true smut" (Bauer et al, 2008). However, some taxa within the Ustilaginomycetes class still represent unresolved complexes. Our phylogenomic tree based on a distance super-matrix approach reinforces that the genera *Ustilago* and *Pseudozyma* are polyphyletic, including non-related species (McTaggart et al., 2012; McTaggart et al., 2016; Wang et al., 2015).

The decomposition of gene trees in quartets shows some phylogenetic conflicts. Dutheil et al. (2016) argues for incomplete lineage sorting as a source of phylogenetic incongruence among smut species genes, but hybridization events can also cause discordant gene tree topologies. Kellner et al. (2011) detected a high potential for hybridization in some extant smut species. Hybridization is recognized as a major force in generate new host specificities (Stukenbrock et al., 2012; Menardo et al., 2016; Depotter et al., 2016) and it may have occurred in the evolutionary history of smut fungi.

4.2) Genomic features of smut pathogens

In this study, we broadly extended the comparative genomic studies on true smut fungi by using all genomes available to date. We additionally sequenced two other genomes from isolates infecting distinct hosts (*Ustilago hordei* from oats and *Ustilago tritici* from wheat) in order to identify unique genomic features that may are responsible for their host specialization.

In several species of filamentous plant pathogens is observed a trend towards genome increasing and proliferation of repetitive elements (Raffaele and Kamoun, 2012). However, the smut pathogens sequenced so far have compact genomes depleted of paralogs and repetitive DNA. *U. hordei* is an exception and seems to be experiencing genome expansion by the prediction of more protein-coding genes and repetitive elements. Dutheil et al. (2016) speculate that the activity of transposons in *U. hordei* genome may are under less stringent control and caused the ungrouping of some candidate effectors by their linked movement throughout the genome. The sequencing of a second *U. hordei* strain herein supports this hypothesis by showing an even greater content of repetitive DNA than the previously sequenced strain. In many cases, the genomic plasticity and rapid evolution of pathogens have been associated with the activity of transposable elements (Wöstemeyer and Kreibich, 2002;

Raffaele and Kamoun, 2012; Castanera et al., 2016). A population genomics study of *U. hordei* has the potential to show the ongoing genomic evolution via transposable elements.

The evolutionary similarities in cell biology and lifestyle among the nine smut species is reflected in the sharing of more than 65% of orthologs groups, most detected as one-to-one orthologs. However, among those genes we found significant evidence of episodic positive selection.

Due to the key role of cell-wall degrading enzymes (CAZymes, proteases and lipases) and effectors during plant-pathogen interactions, we focused on differences in gene content of these categories. Orphan and positively selected genes were also explored in more detail.

4.3) Similar content of plant cell-wall degrading enzymes

Plant cell wall-degrading enzymes play central roles in host penetration and nutrient acquisition during fungal infections. The arsenal of those enzymes varies among fungi, reflecting their lifestyles and host preferences (King et al., 2011; Zhao et al., 2014). Hence, we investigate if there was differential distribution of CAZymes, proteases and lipases categories among the host-specialized smut species. However, the nine smut pathogens showed similar enzyme content. The amount of CAZymes in smut species is in agreement to what is reported for other biotrophic fungi (Zhao et al., 2014). Biotrophic fungi tend to have fewer CAZymes than hemibiotrophs and necrotrophs, causing minimal damages to the hosts (Kim et al., 2016). As reported for biotrophic fungi, smut fungi also lack the glycoside hydrolase family 6 which has a well-known cellulase activity for plant cell wall degradation (Zhao et al., 2014). However, other genes encoding redundant cellulose, hemicellulose, pectin, and cutin degrading enzyme families are present in smut genomes.

Zhao et al. (2014) found that pathogens of dicots often contain more pectinases than fungi infecting monocots. Dicots and noncommelid monocots have “type I” primary cell walls generally enriched in pectin and xyloglucan; while commelinid monocots, such as species from Poales order, contain lower content of pectin and an increased content of glucuroarabinoxylan. Such expansion in pectinase content was not detected in *M. pennsylvanicum* in relation to Poaceae-smut pathogens. However, *M. pennsylvanicum* has species-specific and positively selected CAZymes and proteases that may contribute for the dicot-host adaptation.

The most discrepant pattern among the analyzed enzymes was in the Aspartic peptidase A11A family expanded only in *U. hordei* genomes. A11A family contains

endopepetidases encoded by retrotransposons that act on polyprotein processing (http://merops.sanger.ac.uk). This is another evidence of genome expansion via transposable elements in *U. hordei* species.

4.4) The acquisition of an optimal effector gene repertoire

To complete their lifecycles into living host tissues, biotrophic pathogens harbor a repertoire of secreted effector proteins to repress host defenses and manipulate host physiology to their own purposes (Kemen et al., 2015). By using a machine learning approach, the number of predicted effector genes was highly variable among smut species. The smallest effector repertoire of *M. pennsylvanicum* was already identified by Sharma et al. (2014) who proposed that gene loss is the hallmark of the host jump event to a dicot host.

Most effector proteins lack functional domains and few have been functionally characterized in smut pathogens (Redkar et al., 2015). Among those known effectors, some are present in all smut species and might are core pathogenicity factors for the establishment of the disease or enhancing the fitness. The effector Cmu1 (*Chorismate mutase-1*) is secreted by *U. maydis* into the maize cells and it is required for full virulence. Cmu1 diverts the host chorismate metabolism to the phenylpropanoid pathway thereby outcompeting the synthesis of salicylic acid (Djmaei et al., 2011). The suppression of salicylic acid levels into the host plant likely affects the downstream defense responses, which are particularly important against biotrophic pathogens (Glazebrook, 2005). The effector Stp1 (*stop after penetration 1*) is also essential for complete virulence in *U. maydis*. Stp1 probably acts in papain-like cysteine proteases, blocking its secretion in the plant cytosol and inhibiting its action in the plant apoplast (Liang, 2012). Stp1 may also alter the regulation of transcription in the host nucleus. Many studies have shown the importance of papain-like cysteine proteases as a basal defense mechanism of plants against pathogens and a variety of unrelated pathogen-derived effectors targeting subfamilies of such proteases (Misas-Villamil et al., 2016). The effector ApB73 (*Apathogenic in B73*) of *U. maydis* is required for the successful colonization of a specific maize cultivar B73. ApB73 may have a shielding function, but its mechanistic role remains to be elucidated (Stirnberg and Djamei, 2016). Except for *S. reilianum* ortholog, the cross-species complementation assays in *U. maydis* were unable to rescue the virulence defect of the *apB73* deletion strain (Stirnberg and Djamei, 2016). ApB73 protein from *S. reilianum* and *U. maydis* present the highest percentage of identity and both species infect maize, suggesting that ApB73 may evolve to interact/protect from host specific defenses. Members

of the *eff1* (*effector family 1*) are also present in all smut species, although in distinct numbers. In *U. maydis*, deletion of the entire gene family showed reduction in virulence, but the mutant is still able to establish biotrophic growth (Khrunyk et al., 2010). Despite evidences for functional redundancy of *eff1* members, *eff1-11*, *eff1-3* and *eff1-4* showed stronger contribution to virulence and tumor formation in *U. maydis* than other family members (Khrunyk et al., 2010). The contribution of each *eff1* member and the copy number to the virulence can be different in each species and their molecular function also remains to be investigated.

Other known effectors are not present in all species. Based on the genome searches, distinct evolutionary processes seem to be involved, such as pseudogenization and gene gain/loss. The effector Pep1 (*Protein essential during penetration-1*) is essential for *U. maydis*-maize compatible interaction. Pep1 is an inhibitor of plant peroxidases, acting in oxidative burst suppression (Hemetsberger et al., 2012). The oxidative burst, through the production of reactive oxygen species, is one of the earliest responses of plants upon pathogen attack. Pep1 is absent only in *U. esculenta*. Although a homologous region has been found in the *U. esculenta* genome, a proper start codon for Pep1 does not seem to exist. It is worth mentioning that the *U. esculenta* interaction with *Z. latifolia* results in an edible smut gall and prevents inflorescence. For several centuries, the infected *Z. latifolia* has been cultivated in Taiwan and Southern China. *U. esculenta* spends its entire life into the host plant, being propagated together with the host through asexual rhizome (Chung and Tzeng, 2004). Relaxed selective pressure for penetration and suppression of early defense responses may lead to Pep1 pseudogenization.

See1 (*Seedling efficient effector 1*) is an *U. maydis* organ-specific effector required for tumor formation in vegetative tissues (Schilling et al., 2014). See1 is translocated into the plant cell cytoplasm and nucleus, where it interacts with SGT1 protein inhibiting its MAPK-triggered phosphorylation (Redkar et al., 2015a). Hijacking of SGT1 seems to contribute to the deactivation of salicylic acid-mediated defense responses and reactivation of plant DNA synthesis and host cell cycle during leaf tumor progression. *U. maydis* is an exception among smut pathogen by its ability to infect all aerial parts of the host plant and locally induce tumor formation. See1 might be important for host adaptation of *U. maydis* and the independence of inflorescences may also be an advantage as it allows multiple disease cycles in less time. However, SGT1 is also target by several bacterial effectors as plant defense suppressors (Redkar et al., 2015a) and See1 orthologs are present in other smut genomes. Nonetheless,

Redkar et al. (2015b) showed that the *U. hordei* See1 does not functionally complement the deletion mutant of *U. maydis*. Other candidate leaf-induced effector genes (*um06223* and *um12217*) that also affect tumor formation (Schilling et al., 2014) were specific of *U. maydis* genome, suggesting a role in host adaptation and local symptom development.

The effector Pit2 (*Protein involved in tumors 2*) of *U. maydis* interacts with apoplastic cysteine proteases inhibiting the enzymatic activity. Deletion mutants can infect maize, but elicit defense responses and fail to induce tumors (Mueller et al, 2013). The gene is absent in *U.esculenta* and *U. trichophora* and may have a redundant function with Stp1 effector, as both act on the same line of plant defense. The effector Tin2 (*Tumor inducing 2*) from *U. maydis* was shown to be responsible for anthocyanin induction in maize (Brefort et al., 2014). Tin2 avoids the ubiquitin-proteasome degradation of a protein kinase. The stabilization of the kinase seems to divert the flux of precursors from lignin biosynthetic pathway to anthocyanin biosynthesis (Tanaka et al. 2014). The deposition of lignin in the plant cell wall is a physical barrier to pathogen infection. The *tin2* gene is absent in *M. pennsylvanicum* and both *U. hordei* species. No reports of anthocyanin accumulation for these pathosystems were found in the literature and Tin2 was shown to weakly contribute to virulence of *U. maydis*.

Mig1 and Mig2 (*maize-induced gene*) families are induced during biotrophic growth. The deletion of one *mig1* or the entire *mig2* family did not affect the pathogenicity. The deletion of *mig1* cluster caused hypervirulence phenotype during *U. maydis* infection (Basse et al., 2000; Basse et al., 2002; Schirawski et al., 2010). Mig1 and 2 families are absent in *M. pennsylvanicum* and *U. esculenta* genomes. Mig2 family is also absent in *S. reilianum* and *S. scitamineum*.

Vag2 contribute mostly for spore formation and partially to suppress leaf tip death on *S. reilianum*-maize interaction, showing high expression in ears (Zhao, 2015). Vag2 could interact with several plant proteins, including primary metabolic pathways, binding to DNA or RNA, and signal transduction. Vag2 is absent in *U. tritici*, *M. pennsylvanicum*, *U. esculenta* and both *U. hordei* species.

SAD1 (*Supressor of Apical Dominance 1*) plays a role in regulation of apical dominance during *S. reilianum*-maize interaction altering the branching architecture and increasing the number of ears (Ghareeb et al., 2015). Once inside the plant cell, SAD1 accumulates in the nucleus and could affect apical dominance by altering the gene transcription program. Auxin efflux transporter PIN1 was up-regulated in the maize roots and the suppressor of axillary bud outgrowth TB1 was down-regulated in the stalks upon *S. reilianum* infection (Ghareeb et al., 2015). Ortholog analysis showed that SAD1 is species-

specific, but similarity was found in other genomes. The symptom of loss of apical dominance in *S. scitamineum* and *U. esculenta* is observed as increasing tillering in sugarcane and *Z. latifolia*, respectively (Sundar et al., 2012; Yan et al., 2013). Increased tillering also occurs on *S. reilianum* infecting sorghum (Ghareeb et al. 2015). A detailed analysis of this gene in *S. scitamineum* is being conducted and indicates that, in this case, there was an automatic prediction error.

The *U. hordei* effector UhAvr1 is needed very early during infection. UhAvr1 does not contribute significantly to virulence. However, in barley cultivars harboring the resistance gene *ruh1*, the UhAvr1 presence causes immediate necrotic reaction, resulting in complete immunity (Ali et al., 2014). Orthologs of UhAvr1 are also present in the oats-infecting *U. hordei* and *U. maydis*. However, no effector-R gene relationship have been identified in other smut pathosystems to date (Ali et al., 2014), suggesting that UhAvr1 is highly adapted to *U. hordei*-barley interaction. If UhAvr1 induces the same hypersensitive response in the oats host remains to be studied.

Most of characterized smut effectors are not essential for infection and are not conserved in all species. It is seems that each effector has minor effects on virulence and functional redundancy compensates an effector loss. Moreover, effectors are subject to rapid evolutionary changes and to strong selection pressure to evade the coevolving plant defense mechanism (Sperschneider et al., 2015). Therefore, smut effectors probably had distinct evolutionary fates to adapt to different hosts and some effectors may have diverged to an extent that they are no longer recognized as orthologs by our criteria. Diversifying selection was observed for few one-to-one orthologs groups harboring candidate effectors, but sets of lineage-specific orphan effectors were detected by our comparative study. We suggest that these positively selected and orphan effector genes are good candidates for further characterization in regards to their role in virulence and host specificity.

4.5) Orphan and positively selected genes: potential metabolic versatility, microbiota interaction, and host molecule recognition

By adding an increased number of species in this comparative genomics study, we found a smaller number of orphan genes than in the comparison of four genomes performed by Sharma et al. (2014) and Taniguti et al. (2015), indicating a reduced number of false positives. The majority of the orphan genes in all smut species have no functional domain assigned challenging biological insights. Despite of presenting the lowest content of secreted proteins and predicted effectors among smut species, *M. pennsylvanicum* has the highest

number of genes under positive selection.

Some orphan and positively selected genes were associated with primary and secondary metabolic pathways, suggesting metabolic versatility among smut fungi. Wisecaver et al. (2014) analyzing Enzyme Commission (EC)-annotated metabolic genes for 208 fungal genomes found that gene duplication followed by diversification is the dominant process underlying fungal metabolic diversity. Interestingly, *S. reilianum* has enriched orphan enzymes acting on oxidative phosphorylation pathway (reductases, oxidases, and dehydrogenases). Such enzymes participate in energy production, but also in antioxidant defense of fungal cells (Marcet-Houben et al.,2009). Marcet-Houben et al. (2009) have shown that gene loss and gene duplication followed by functional divergence have shaped the respiratory pathway in many fungal species and may have allow the adaptation to different life-styles and environments. Another interesting gene considered specific of *M. pennsylvanicum* genome is a putative secreted protein with Pfam domain for tannase/feruloyl esterase. This enzyme is involved in polyphenol degradation. Different tannases act on different plant polyphenols, which may suggest an adaptation to the host biochemical composition.

Secondary metabolites, as toxins, were shown previously to play important roles in virulence and host specificity of many Ascomycota plant pathogens (Ohm et al., 2012). Few cluster of genes encoding enzymes for secondary metabolite production are present in smut genomes and no phytotoxin production by smut fungi is reported in the literature. The diversification of secondary metabolite enzymes in smut fungi may are associated with the production of diverse signaling molecules, sporogenic factors, or antimicrobial agents.

Orphan and positively selected genes associated with response to stimulus, signaling, transduction, regulation of transcription, morphogenesis, and cell-cycle programming may are related to the orchestration of infection and development of each smut pathogen into each host. Smut species present distinct sporulation time and penetrate at distinct host sites. It is likely that different host molecules are perceived as a signal for penetration and induction of fungal sporogenesis. Smut species preferably penetrates through the coleoptile of young seedlings, *S. reilianum* can penetrate through the maize roots, *S. scitamineum* through young sugarcane buds, and *U. maydis* through all maize aerial parts. The sporogenesis of most smut fungi occurs after flowering, replacing the grain kernel of infected plants by masses of teliospores. In *S. reilianum*, sporogenesis is concomitant with floral initiation of the maize and changes in glycoprotein content of cell walls following transition from vegetative to floral growth was associated with the induction of fungal sporogenesis (Gahreeb, 2011). The

sporogenesis in *S. scitamineum* and *U. esculenta* occurs beneath the apical meristem and prevents their host to produce inflorescences. However, in *S. scitamineum* the sporogenesis culminates with the out-growth of a whip-like structure in the sugarcane apex (similar to flowering structure) (Schaker et al., 2016; Marques et al., 2016); while in *U. esculenta* the teliospores remains inside the galls into the node region (Chung and Tzeng, 2004). The sporogenesis of *U. maydis* can occur at maize leaves, independent of flowering transitions (Brefort et al., 2009). Further studies are required to test all hypotheses raised herein.

Using site-specific models, positive selection acting on specific codons were found for 31 genes. Among those, we identified positively selected sites in mating-type *bE* gene. Mating genes are shown to evolve faster than other genes in *Neurospora* species and positive selection was found in heterothallic taxa (Wik et al, 2008). In smut fungi, *bE* locus encoding for a component of the heterodimeric bE/bW homeodomain transcription factor that triggers filamentous growth and pathogenicity after compatible yeast-like cell recognition and fusion (Bakkeren et al., 2008). Positive selection at *bE* sites can promote reproductive isolation among species by non-dimerization with *bW* or maybe selection was identified just due to bias in allele sampling in the sequenced genomes.

Specific-sites selection was also identified in a putative diacylglycerol acyltransferase enzyme that acts in the final step of triacylglycerol synthesis (TG). TG is a storage lipid which serves as energy reservoir, source of signaling molecules, and substrate for membrane biogenesis (Liu et al., 2012). The TG biosynthesis pathway is conserved in all living organisms; however sequence motifs of diacylglycerol acyltransferase are not conserved (Turchetto-Zolet et al., 2011). In *S. scitamineum*, the gene encoding for this enzyme was upregulated during sporogenesis stage (Taniguti et al., 2015) which may be related to the accumulation of lipid droplets in teliospores that will serve as a source of energy during germination (Marques et al., 2017). This enzyme was also associated with pathogenicity in the broad host range pathogen *Colletotrichum gloeosporioides* (Sharma et al., 2016). The significance of the selected sites for functional differences remains to be explored.

5) CONCLUSION

Our comparative genomic study provided insights on smut pathogen biology and symptoms development. Complex genetic basis underlies host specialization in smut fungi and may involve series of episodic selection events at shared genes and gene loss/gain events generating species-specific genes. However, it is difficult to distinguish the genetic changes that directly contribute to the host specialization from those that were a consequence of the

divergence after host specialization. Various metabolic pathways, regulation of transcription, and signaling processes were affected in our analyses. We emphasize the acquisition and maintenance of an optimal effector repertoire as the main determinants of host specificity. Further functional studies are required to test this hypothesis. It is also possible that the different host environments affect the expression pattern and functionality of effectors genes, which also remains to be investigated. Population genomics studies of *U. hordei* would also be interesting to reveal the ongoing genome evolution via transposable elements and the transposition frequency.

REFERENCES

Albert H. H., Schenck S., 1996 PCR amplification from a homolog of the bE mating-type gene as a sensitive assay for the presence of *Ustilago scitaminea* DNA. Plant Dis. **80**: 1189–1192.

Ali S., Laurie J. D., Linning R., Cervantes-Chávez J. A., Gaudet D., Bakkeren G., 2014 An Immunity-Triggering Effector from the Barley Smut Fungus *Ustilago hordei* Resides in an Ustilaginaceae-Specific Cluster Bearing Signs of Transposable Element-Assisted Evolution. PLOS Pathog. **10**: e1004223.

Antonovics J., Boots M., Ebert D., Koskella B., Poss M., Sadd B. M., 2013 The origin of specificity by means of natural selection: Evolved and nonhost resistance in host-pathogen interactions. Evolution (N. Y). **67**: 1–9.

Bakkeren G., Kämper J., Schirawski J., 2008 Sex in smut fungi: Structure, function and evolution of mating-type complexes. Fungal Genet. Biol. **45**.

Baltrus D. A., Nishimura M. T., Dougherty K. M., Biswas S., Mukhtar M. S., Vicente J., Holub E. B., Dangl J. L., 2012 The Molecular Basis of Host Specialization in Bean Pathovars of *Pseudomonas syringae*. Mol. Plant-Microbe Interact. **25**: 877–888.

Barrett L. G., Kniskern J. M., Bodenhausen N., Zhang W., Bergelson J., 2009 Continua of specificity and virulence in plant host-pathogen interactions: causes and consequences. New Phytol. **183**: 513–529.

Barrett L. G., Heil M., 2012 Unifying concepts and mechanisms in the specificity of plant–enemy interactions. Trends Plant Sci. **17**: 282–292.

Basse C. W., Stumpferl S., Kahmann R., 2000 Characterization of a *Ustilago maydis* Gene Specifically Induced during the Biotrophic Phase: Evidence for Negative as Well as Positive Regulation. Mol. Cell. Biol. **20**: 329–339.

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

Bauer R., Begerow D., Franz O., 2008 The true smut fungi. Accessed: http://tolweb.org/Ustilaginomycotina/20530/2008.01.23.

Begerow D., Goker M., Lutz M., Stoll M., 2004 On the evolution of smut fungi on their hosts. In: Agerer R, Piepenbring M, Blanz PA (Eds.), *Frontiers in Basidiomycote Mycology*, IHW-Verlag, Eching, Germany, pp. 81–98.

Bendtsen J. D., Jensen L. J., Blom N., Heijne G. von, Brunak S., 2004 Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. **17**: 349–356.

Bhat T., Singh B., Sharma O., 1998 Microbial degradation of tannins – A current perspective. Biodegradation **9**: 343–357.

Bölker M., 2001 *Ustilago maydis*--a valuable model system for the study of fungal dimorphism and virulence. Microbiology **147**: 1395–1401.

Brefort T., Tanaka S., Neidig N., Doehlemann G., Vincon V., Kahmann R., 2014 Characterization of the Largest Effector Gene Cluster of *Ustilago maydis*. PLOS Pathog. **10**: e1003866.

Burmester A., Shelest E., Glöckner G., Heddergott C., Schindler S., Staib P., *et al.*, 2011 Comparative

and functional genomics provide insights into the pathogenicity of dermatophytic fungi. Genome Biol. **12**: R7.

Castanera R., López-Varas L., Borgognone A., LaButti K., Lapidus A., Schmutz J., Grimwood J., Pérez G., Pisabarro A. G., Grigoriev I. V, Stajich J. E., Ramírez L., 2016 Transposable Elements *versus* the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. PLoS Genet **12**: e1006108.

Castresana J., 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17**: 540–552.

Chin C.-S., Alexander D. H., Marks P., Klammer A. a, Drake J., Heiner C., Clum A., Copeland A., Huddleston J., Eichler E. E., Turner S. W., Korlach J., 2013 Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods **10**: 563–569.

Choi Y.-J., Thines M., 2015 Host Jumps and Radiation, Not Co-Divergence Drives Diversification of Obligate Pathogens. A Case Study in Downy Mildews and Asteraceae (M Gijzen, Ed.). PLoS One **10**: e0133655.

Chung K.-R., Tzeng D., 2004 Nutritional requirements of the edible gall-producing fungus *Ustilago esculenta*. J. Biol. Sci. **4**: 246–252.

Conesa A., Götz S., García-Gómez J. M., Terol J., Talón M., Robles M., 2005 Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**: 3674–3676.

Darriba D., Taboada G. L., Doallo R., Posada D., 2011 ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics **27**: 1164–1165.

Depotter J. R. L., Seidl M. F., Wood T. A., Thomma B. P. H. J., 2016 Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. Curr. Opin. Microbiol. **32**: 7–13.

Dilokpimol A., Mäkelä M. R., Aguilar-Pontes M. V., Benoit-Gelber I., Hildén K. S., Vries R. P. de, 2016 Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. Biotechnol. Biofuels **9**: 231.

Djamei A., Schipper K., Rabe F., Ghosh A., Vincon V., Kahnt J., *et al.*, 2011 Metabolic priming by a secreted fungal effector. Nature **478**: 395–398.

Dutheil J. Y., Mannhaupt G., Schweizer G., Sieber C. M. K., Münsterkötter M., Güldener U., Schirawski J., Kahmann R., 2016 A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi. Genome Biol. Evol. .

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

English A. C., Salerno W. J., Reid J. G., 2014 PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. BMC Bioinformatics **15**: 1–7.

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

Fischer M., Pleiss J., 2003 The Lipase Engineering Database: a navigation and analysis tool for protein families. Nucleic Acids Res. **31**: 319–321.

Ghareeb H., 2011 Molecular Dissection of Maize-*Sporisorium reilianum* Interactions: Host Developmental Changes and Pathogen Effectors. Dissertation from Max Planck Institute.

Ghareeb H., Drechsler F., Löfke C., Teichmann T., Schirawski J., 2015 SUPPRESSOR OF APICAL DOMINANCE1 of *Sporisorium reilianum* Modulates Inflorescence Branching Architecture in Maize and Arabidopsis. Plant Physiol. **169**: 2789–2804.

Gilbert G. S., Webb C. O., 2007 Phylogenetic signal in plant pathogen-host range. Proc. Natl. Acad. Sci. U. S. A. **104**: 4979–4983.

Gioti A., Nystedt B., Li W., Xu J., Andersson A., Averette A. F., *et al.*, 2013 Genomic Insights into the Atopic Eczema-Associated Skin Commensal Yeast *Malassezia sympodialis*. MBio **4**: e00572-12.

Giraud T., Refrégier G., Gac M. Le, Vienne D. M. de, Hood M. E., 2008 Speciation in fungi. Fungal Genet. Biol. **45**: 791–802.

Giraud T., Gladieux P., Gavrilets S., 2010 Linking emergence of fungal plant diseases and ecological speciation. Trends Ecol. Evol. **25**: 387–395.

Glazebrook J., 2005 Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic

Pathogens. Annu. Rev. Phytopathol. **43**: 205–227.

Grandaubert J., Bhattacharyya A., Stukenbrock E. H., 2015 RNA-seq-Based Gene Annotation and Comparative Genomics of Four Fungal Grass Pathogens in the Genus *Zymoseptoria* Identify Novel Orphan Genes and Species-Specific Invasions of Transposable Elements. G3 Genes|Genomes|Genetics **5**: 1323 LP-1333.

Haueisen J., Stukenbrock E. H., 2016 Life cycle specialization of filamentous pathogens — colonization and reproduction in plant tissues. Curr. Opin. Microbiol. **32**: 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 Pathog. **8**.

Huerta-Cepas J., Serra F., Bork P., 2016 ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. **33**: 1635–1638.

Johnson K. P., Malenke J. R., Clayton D. H., 2009 Competition promotes the evolution of host generalists in obligate parasites. Proc. R. Soc. B Biol. Sci. **276**: 3921–3926.

Jones P., Binns D., Chang H.-Y., Fraser M., Li W., McAnulla C., *et al.*, 2014 InterProScan 5: genome-scale protein function classification. Bioinforma. **30**: 1236–1240.

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., Vollmeister E., Feldbrügge M., Begerow D., 2011 Interspecific Sex in Grass Smuts and the Genetic Diversity of Their Pheromone-Receptor System. PLoS Genet. **7**: e1002436.

Kemen A. C., Agler M. T., Kemen E., 2015 Host–microbe and microbe–microbe interactions in the evolution of obligate plant parasitism. New Phytol. **206**: 1207–1228.

Khrunyk Y., Münch K., Schipper K., Lupas A. N., Kahmann R., 2010 The use of FLP-mediated recombination for the functional analysis of an effector gene family in the biotrophic smut fungus *Ustilago maydis*. New Phytol. **187**: 957–968.

Kim K.-T., Jeon J., Choi J., Cheong K., Song H., Choi G., Kang S., Lee Y.-H., 2016 Kingdom-wide analysis of fungal small secreted proteins (SSPs) reveals their potential role in host association. Front. Plant Sci. **7**.

King B. C., Waxman K. D., Nenni N. V, Walker L. P., Bergstrom G. C., Gibson D. M., 2011 Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. Biotechnol. Biofuels **4**: 4.

Kirzinger M. W. B., Stavrinides J., 2012 Host specificity determinants as a genetic continuum. Trends Microbiol. **20**: 88–93.

Konishi M., Hatada Y., Horiuchi J., 2013 Draft Genome Sequence of the Basidiomycetous Yeast-Like Fungus *Pseudozyma hubeiensis* SY62, Which Produces an Abundant Amount of the Biosurfactant Mannosylerythritol Lipids. Genome Announc. **1**: e00409-13.

Laurie J. D., Ali S., Linning R., Mannhaupt G., Wong P., Güldener U., Münsterkötter M., Moore R., Kahmann R., Bakkeren G., Schirawski J., 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**: 1733–45.

Lefebvre F., Joly D. L., Labbé C., Teichmann B., Linning R., Belzile F., Bakkeren G., Bélanger R. R., 2013 The Transition from a Phytopathogenic Smut Ancestor to an Anamorphic Biocontrol Agent Deciphered by Comparative Whole-Genome Analysis. Plant Cell **25**: 1946–1959.

Li L., Stoeckert C. J., Roos D. S., 2003 OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. **13**: 2178–2189.

Liang L., 2012 The role of Stp1, a secreted effector, in the biotrophic interaction of *Ustilago maydis* and its host plant maize.Dissertation from Philipps-Universität Marburg.

Liu Q., Siloto R. M. P., Lehner R., Stone S. J., Weselake R. J., 2012 Acyl-CoA:diacylglycerol acyltransferase: Molecular biology, biochemistry and biotechnology. Prog. Lipid Res. **51**: 350–377.

Lombard V., Golaconda Ramulu H., Drula E., Coutinho P. M., Henrissat B., 2013 The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res **42**.

Lorenz S., Guenther M., Grumaz C., Rupp S., Zibek S., Sohn K., 2014 Genome Sequence of the Basidiomycetous Fungus *Pseudozyma aphidis* DSM70725, an Efficient Producer of Biosurfactant Mannosylerythritol Lipids. Genome Announc. **2**: e00053-14.

Ma L.-J., Does H. C. van der, Borkovich K. A., Coleman J. J., Daboussi M.-J., Pietro A. Di,*et al.*, 2010 Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium.* Nature **464**: 367–373.

Marcet-Houben M., Marceddu G., Gabaldón T., 2009 Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. BMC Evol. Biol. **9**: 295.

Marques J. P. R., Appezzato-da-Glória B., Piepenbring M., Massola Jr N. S., Monteiro-Vitorello C. B., Vieira M. L. C., 2017 Sugarcane smut: shedding light on the development of the whip-shaped sorus. Ann. Bot. **119**: 815–827.

Matei A., Doehlemann G., 2016 Cell biology of corn smut disease — *Ustilago maydis* as a model for biotrophic interactions. Curr. Opin. Microbiol. **34**: 60–66.

Matheussen A. M., Morgan P. W., Frederiksen R. A., 1991 Implication of gibberellins in head smut (*Sporisorium reilianum*) of Sorghum bicolor. Plant Physiol **96**.

McTaggart A. R., Shivas R. G., Geering A. D. W., Callaghan B., Vánky K., Scharaschkin T., 2012 Soral synapomorphies are significant for the systematics of the *Ustilago-Sporisorium-Macalpinomyces* complex (Ustilaginaceae). Persoonia Mol. Phylogeny Evol. Fungi **29**: 63–77.

McTaggart A. R., Shivas R. G., Boekhout T., Oberwinkler F., Vánky K., Pennycook S. R., Begerow D., 2016 *Mycosarcoma* (Ustilaginaceae), a resurrected generic name for corn smut (*Ustilago maydis*) and its close relatives with hypertrophied, tubular sori. IMA Fungus **7**: 309–315.

Medema M. H., Blin K., Cimermancic P., Jager V. de, Zakrzewski P., Fischbach M. A., Weber T., Takano E., Breitling R., 2011 antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res **39**.

Menardo F., Praz C. R., Wyder S., Ben-David R., Bourras S., Matsumae H., *et al.*, 2016 Hybridization of powdery mildew strains gives rise to pathogens on novel agricultural crop species. Nat Genet **48**: 201–205.

Misas-Villamil J. C., Hoorn R. A. L. van der, Doehlemann G., 2016 Papain-like cysteine proteases as hubs in plant immunity. New Phytol. **212**: 902–907.

Morita T., Koike H., Hagiwara H., Ito E., Machida M., Sato S., Habe H., Kitamoto D., 2014 Genome and Transcriptome Analysis of the Basidiomycetous Yeast *Pseudozyma antarctica* Producing Extracellular Glycolipids, Mannosylerythritol Lipids. PLoS One **9**: e86490.

Morrow C. A., Fraser J. A., 2009 Sexual reproduction and dimorphism in the pathogenic basidiomycetes. FEMS Yeast Res. **9**: 161–177.

Mueller A. N., Ziemann S., Treitschke S., Aßmann D., Doehlemann G., 2013 Compatibility in the *Ustilago maydis*–Maize Interaction Requires Inhibition of Host Cysteine Proteases by the Fungal Effector Pit2. PLOS Pathog. **9**: e1003177.

Munkacsi A. B., Stoxen S., May G., 2007 Domestication of maize, sorghum, and sugarcane did not drive the divergence of their smut pathogens. Evolution **61**: 388–403.

Ohm R. A., Feau N., Henrissat B., Schoch C. L., Horwitz B. A., Barry K. W., *et al.*,, 2012 Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen Dothideomycetes Fungi. PLOS Pathog. **8**: e1003037.

Oliveira J. V. de C., Santos R. A. C. dos, Borges T. A., Riaño-Pachón D. M., Goldman G. H., 2013 Draft Genome Sequence of *Pseudozyma brasiliensis sp*. nov. Strain GHG001, a High Producer of Endo-1,4-Xylanase Isolated from an Insect Pest of Sugarcane. Genome Announc. **1**: e00920-13.

Piepenbring M., 2009 Diversity, ecology and systematics of smut fungi. In: Claro K Del, Oliveira P, Rico-Gray V (Eds.), *Tropical Biology and Conservation Management*, EOLSS Publications, p. .

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. Front. Plant Sci. **8**: 119.

Poppe S., Dorsheimer L., Happel P., Stukenbrock E. H., 2015 Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen *Zymoseptoria tritici* (*Mycosphaerella graminicola*). PLoS Pathog **11**: e1005055.

Raffaele S., Farrer R. A., Cano L. M., Studholme D. J., MacLean D., Thines M., 2010 Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science **330** :1540-1543.

Raffaele S., Kamoun S., 2012 Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Micro **10**: 417–430.

Rawlings N. D., Barrett A. J., Bateman A., 2012 MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. **40**: 343–350.

Redkar A., Hoser R., Schilling L., Zechmann B., Krzymowska M., Walbot V., Doehlemann G., 2015a A Secreted Effector Protein of *Ustilago maydis* Guides Maize Leaf Cells to Form Tumors. Plant Cell **27**(4):1332-1351.

Redkar A., Villajuana-Bonequi M., Doehlemann G., 2015b Conservation of the *Ustilago maydis* effector See1 in related smuts. Plant Signal. Behav. **10**: e1086855.

Sayyari E., Mirarab S., 2016 Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Mol. Biol. Evol. **33**: 1654–1668.

Schaker P. D. C., Palhares A. C., Taniguti L. M., Peters L. P., Creste S., Aitken K. S., Sluys M.-A. Van, Kitajima J. P., Vieira M. L. C., Monteiro-Vitorello C. B., 2016 RNAseq Transcriptional Profiling following Whip Development in Sugarcane Smut Disease. PLoS One **11**: e0162237.

Schilling L., Matei A., Redkar A., Walbot V., Doehlemann G., 2014 Virulence of the maize smut Ustilago maydis is shaped by organ-specific effectors. Mol. Plant Pathol. **15**: 780–789.

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**: 1546–1548.

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 Melanopsichium pennsylvanicum. Genome Biol Evol **6**: 2034–2049.

Sharma M., Guleria S., Kulshrestha S., 2016 Diacylglycerol acyl transferase: A pathogenicity related gene in *Colletotrichum gloeosporioides*. J. Basic Microbiol. **56**: 1308–1315.

Slater G. S. C., Birney E., 2005 Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6**.

Sperschneider J., Dodds P. N., Gardiner D. M., Manners J. M., Singh K. B., Taylor J. M., 2015 Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. PLOS Pathog. **11**: e1004806.

Sperschneider J., Gardiner D. M., Dodds P. N., Tini F., Covarelli L., Singh K. B., Manners J. M., Taylor J. M., 2016 EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytol. **210**: 743–761.

Stamatakis A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**: 1312–1313.

Stanke M., Morgenstern B., 2005 AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. **33**: 465–467.

Stirnberg A., Djamei A., 2016 Characterization of ApB73, a virulence factor important for colonization of *Zea mays* by the smut *Ustilago maydis* . Mol. Plant Pathol. **17**: 1467–1479.

Stukenbrock E. H., Christiansen F. B., Hansen T. T., Dutheil J. Y., Schierup M. H., 2012 Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc. Natl. Acad. Sci. **109**: 10954–10959.

Sundar A. R., Barnabas E. L., Malathi P., Viswanathan R., 2012 A Mini-Review on Smut Disease of Sugarcane Caused by *Sporisorium scitamineum*. In: Mworia JK (Ed.), *Botany*, Intech Open Access Publishers, p. .

Suyama M., Torrents D., Bork P., 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34**: 609–612.

Tanaka S., Brefort T., Neidig N., Djamei A., Kahnt J., Vermerris W., Koenig S., Feussner K., Feussner I., Kahmann R., 2014 A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize. Elife **3**: e01355.

Taniguti L. M., Schaker P. D. C., Benevenuto J., Peters L. P., Carvalho G., Palhares A., *et al.*, 2015 Complete Genome Sequence of *Sporisorium scitamineum* and Biotrophic Interaction Transcriptome with Sugarcane. PLoS One **10**: e0129318.

Toh S. S., Perlin M. H., 2016 Resurgence of Less-Studied Smut Fungi as Models of Phytopathogenesis in the Omics Age. Phytopathology **106**: 1244–1254.

Turchetto-Zolet A. C., Maraschin F. S., Morais G. L. de, Cagliari A., Andrade C. M. B., Margis-Pinheiro M., Margis R., 2011 Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. BMC Evol. Biol. **11**: 263.

Vienne D. M. de, Refrégier G., López-Villavicencio M., Tellier A., Hood M. E., Giraud T., 2013 Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. New Phytol. **198**: 347–385.

Walker B. J., Abeel T., Shea T., Priest M., Abouelliel A., Sakthikumar S., Cuomo C. A., Zeng Q., Wortman J., Young S. K., Earl A. M., 2014 Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One **9**: e112963.

Wang Q.-M., Begerow D., Groenewald M., Liu X.-Z., Theelen B., Bai F.-Y., Boekhout T., 2015 Multigene phylogeny and taxonomic revision of yeasts and related fungi in the Ustilaginomycotina. Stud. Mycol. **81**: 55–83.

Wik L., Karlsson M., Johannesson H., 2008 The evolutionary trajectory of the mating-type (mat) genes in *Neurospora* relates to reproductive behavior of taxa. BMC Evol. Biol. **8**: 109.

Wisecaver J. H., Slot J. C., Rokas A., 2014 The Evolution of Fungal Metabolic Pathways. PLoS Genet. **10**: e1004816.

Wöstemeyer J., Kreibich A., 2002 Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. Curr. Genet. **41**: 189–198.

Xu J., Saunders C. W., Hu P., Grant R. A., Boekhout T., Kuramae E. E., *et al.*, 2007 Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. Proc. Natl. Acad. Sci. **104**: 18730–18735.

Yan N., Wang X.-Q., Xu X.-F., Guo D.-P., Wang Z.-D., Zhang J.-Z., Hyde K. D., Liu H.-L., 2013 Plant growth and photosynthetic performance of *Zizania latifolia* are altered by endophytic *Ustilago esculenta* infection. Physiol. Mol. Plant Pathol. **83**: 75–83.

Yang Z., 2007 PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. **24**: 1586–1591.

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 Res **40**: 445–451.

Yoshida K., Saunders D. G. O., Mitsuoka C., Natsume S., Kosugi S., Saitoh H., Inoue Y., Chuma I., Tosa Y., Cano L. M., Kamoun S., Terauchi R., 2016 Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC Genomics **17**: 1–18.

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**.

Zhao Z., Liu H., Wang C., Xu J.-R., 2014 Erratum to: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics **15**: 1–15.

Zhao Y., 2015 The molecular basis of symptom formation in *Sporisorium reilianum*. Dissertation from RWTH Aachen University.

Zhong Z., Norvienyeku J., Chen M., Bao J., Lin L., Chen L., Lin Y., *et al.*, 2016 Directional Selection from Host Plants Is a Major Force Driving Host Specificity in *Magnaporthe* Species. Sci. Rep. **6**: 25591.

Zmasek C. M., Eddy S. R., 2002 RIO: analyzing proteomes by automated phylogenomics using resampled inference of orthologs. BMC Bioinformatics **3**.

Supplementary Table 1. Specific codons detected under positive selection by M2M1 and M8M7 model comparisons. Genes highlighted in bold were upregulated during late states of the interaction between *S. scitamineum*-sugarcane.

orthogroup	M2M1	M8M7	PFAM	Description
ortho0490	0	0		hypothetical protein UMAG_05145
ortho0617	0	0.00263	TAF6 C-terminal HEAT repeat domain, TATA box binding protein associated factor (TAF)	TATA binding
ortho0644	0.007242	1.50E-05	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family	ribosomal 60S subunit L8
ortho1206	1.80E-05	0.000109	Ribosomal L37ae protein family	probable RPL43B-60S large subunit ribosomal
ortho1351	**0**	**0.024437**	**Diacylglycerol acyltransferase**	**diacylglycerol acyltransferase type 2b**
ortho1813	0	0	Iron permease FTR1 family	high-affinity iron permease
ortho1979	0.029021	0.013901	Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD	peptidyl-prolyl cis-trans partial
ortho2051	2.40E-05	0	Domain of unknown function (DUF543)	conserved hypothetical protein
ortho2262	1.20E-05	0	Ribosomal L38e protein family	ribosomal 60S subunit L38
ortho2306	0	0	Homeobox KN domain, Ustilago B locus mating-type protein	b mating type bE1 allele
ortho2707	5.30E-05	0.000125	Malic enzyme, N-terminal domain, Malic enzyme, NAD binding domain	related to NADP-dependent malic enzyme
ortho2792	0.023547	0.001296	Zinc-binding dehydrogenase, AMP-binding enzyme, Acetyl-coenzyme A synthetase N-terminus, AMP-binding enzyme C-terminal domain	hypothetical protein UMAG_11350
ortho2866	0.000764	1.00E-06	Ribosomal protein L35Ae	ribosomal 60S subunit L33
ortho2871	0.010579	9.80E-05	60s Acidic ribosomal protein	ribosomal P2
ortho3031	0	0	Ribosomal protein S27	probable 40S ribosomal S27
ortho3096	0	0	Elongation factor Tu GTP binding domain, Elongation factor Tu C-terminal domain, Elongation factor Tu domain 2	translation elongation factor EF-1 alpha
ortho3154	0.000274	0.016473	N1221-like protein, Domain of unknown function (DUF3402)	hypothetical protein UMAG_10285
ortho3166	0.00012	0	Helicase conserved C-terminal domain, DEAD/DEAH box helicase	ATP-dependent RNA helicase SUB2
ortho3415	0	0.001236	MmgE/PrpD family	2-methylcitrate dehydratase
ortho3587	0.000103	5.40E-05	Threonyl and Alanyl tRNA synthetase second additional domain, tRNA synthetase class II core domain (G, H, P, S and T), Anticodon binding domain, TGS domain	threonine--tRNA ligase THS1
ortho3614	0	0	KH domain	hypothetical protein UMAG_03654
ortho3630	0.001489	0.003451	Ribosomal protein L3	ribosomal 60S subunit L3
ortho3876	0.018179	0	Ketopantoate hydroxymethyltransferase	likely ketopantoate hydroxymethyltransferase
ortho4192	0.036581	0	rRNA-processing arch domain, DSHCT (NUC185) domain, Helicase conserved C-terminal domain, DEAD/DEAH box helicase	ATP-dependent RNA helicase MTR4
ortho4248	**0.026814**	**0.000347**		**hypothetical protein UMAG_11654**
ortho4268	0.010486	8.30E-05	Ribosomal L29e protein family	probable RPL29-60S large subunit ribosomal
ortho4287	0.00015	0	HMG (high mobility group) box	nonhistone chromosomal
ortho4312	0.005226	0.000664	Ribosomal protein S12/S23	ribosomal 40S subunit S23
ortho4359	**1.00E-06**	**0.049506**		**hypothetical protein UMAG_06152**
ortho4529	0	1.00E-06	Conserved region of unknown function on GLTSCR protein	hypothetical protein UMAG_10422
ortho4704	0.000798	9.00E-06	Ubiquitin-conjugating enzyme	E2 ubiquitin-conjugating enzyme