UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Programa Interunidades de Pós-Graduação em Bioinformática

Gabriel Quintanilha Peixoto

# THE DEATH IS RED: analysis of the predicted secretome of *Aspergillus welwitschiae*, with emphasis in pathogenicity and carbohydrate metabolism

Belo Horizonte

2019

Gabriel Quintanilha Peixoto

# THE DEATH IS RED: analysis of the predicted secretome of *Aspergillus welwitschiae*, with emphasis in pathogenicity and carbohydrate metabolism

**Versão final**

Dissertação apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Bioinformática.

Orientador: Prof. Dr. Aristóteles Góes Neto

Coorientador: Prof. Dr. Eric R. G. R. Aguiar e Profa. Dra. Fernanda Badotti

Belo Horizonte

2019

043 Peixoto, Gabriel Quintanilha.
The death is red: analysis of the predicted secretome of Aspergillus welwitschiae, with emphasis in pathogenicity and carbohydrate metabolism [manuscrito] / Gabriel Quintanilha Peixoto. – 2019.
69 f. : il. ; 29,5 cm.

Orientador: Aristóteles Góes Neto. Co-orientadores: Eric R. G. R. Aguiar e Fernanda Badotti.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Fungos. 3. Aspergillus - patogenicidade. 4. Interações Hospedeiro-Patógeno. I. Góes Neto, Aristóteles. II. Aguiar, Eric Roberto Guimarães Rocha. III. Badotti, Fernanda. IV. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. V. Título

CDU: 573:004

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis – CRB 6 - 2680

*Universidade Federal de Minas Gerais*
*Instituto de Ciências Biológicas*
*Programa Interunidades de Pós-Graduação em Bioinformática da UFMG*

## *ATA DA DEFESA DE DISSERTAÇÃO*

65/2019
entrada
2º/2017
CPF:
134.019.407-40

## Gabriel Quintanilha Peixoto

Às quatorze horas do dia **22 de julho de 2019**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"The Death Is Red: Analysis Of The Predicted Secretome Of Aspergillus Welwitschiae, With Emphasis In Pathogenicity And Carbohydrate Metabolism"**, requisito para obtenção do grau de Mestre em **Bioinformática.** Abrindo a sessão, o Presidente da Comissão, **Dr. Aristóteles Góes Neto**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

| Prof./Pesq. | Instituição | CPF | Indicação |
|---|---|---|---|
| Dr. Aristóteles Góes Neto | UFMG | 549348806-20 | APROVADO |
| Dr. Eric Roberto Guimarães Rocha Aguiar | UFBA | 025021645 08 | APROVADO |
| Dr. Siomar de Castro Soares | UFTM | 05695182611 | APROVADO |
| Dr. Sandeep Tiwari | UFMG | 020.977186-04 | APROVADO |
| Dr. Carlos Priminho Pirovani | UESC | 001 71845765 | APROVADO |

Pelas indicações, o candidato foi considerado: APROVADO
O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.
**Belo Horizonte, 22 de julho de 2019.**

Dr. Aristóteles Góes Neto - Orientador _____

Dr. Eric Roberto Guimarães Rocha Aguiar - Coorientador _____

Dr. Siomar de Castro Soares _____

Dr. Sandeep Tiwari _____

Dr. Carlos Priminho Pirovani _____

*In memoriam of Nelma Alves Quintanilha.*

## ACKNOWLEDGMENTS

There are many people that I would like to thank for the completion of this work, as well as my whole journey in the last two years. But I would like to start by thanking two forces of nature whose guidance and patience brought me to this point; God and my mother, Danielle.

For their contributions to the completion and quality of this work, I would like to thank Dr. Aristóteles Góes Neto, Dr. Vasco Azevedo and Dr. Luiz Goulart, whose efforts made possible for me to continue my studies, my co-advisors Eric Aguiar and Fernanda Badotti for their incredible contributions to this research. To the very agile Sheila Santana, Márcia Natália, Fernanda Magalhães and Tiago, for their availabity to help whenever needed, and special thank Dr. Siomar de Castro Soares, Dr. Sandeep Tiwari, and Dr. Carlos Priminho Pirovani or accepting our invitation and for their contibutions to this work.

Special thanks to Dr. Aline Intorne, for giving me the first chance to work with science as an undergraduate student at the State University of Northern Rio de Janeiro, but also for turning my attention to bioinformatics as a carreer, which led me to pursue my M. Sc. degree in this very area at UFMG.

Thank you to the amazing friends this institution brought me, Gilberto Neto, Juliane Amorim and Luiz Marcelo, for the many conversations and laughs, Raquel Hurtado, Roselane Gomes and Stephane Tosta, for going through this formation process with me and keeping it light, to the LBMCF team including the very talented Dener Bortolini, the very promising undergrads Daniel Araújo and Ruth de Paula, to Luz Alba, to Elizabeth Duarte, Rodrigo Kato, Doglas and Mariana Parise for their help and contributions to this research, and special thanks to my *partners in crime* Luiz Marcelo and Paula Luize, for the best conversations, advices, laughs and or being the best coworkers in the entire world.

To my long-distance friends, Ally Te, Anderson Ribeiro, Daniel Monteiro, Sonia Guimarães, and Vitor Cyrino, for sending good energies from a far. To my partner Gabriel Bruno, for keeping up with my complaints and stress with neverending patience and love.

And finally, to all the people that made this work possible, directly or not, thank you very much!

## ABSTRACT (in Portuguese)

Em 2018, *Aspergillus welwitschiae* foi descrito como causador da podridão vermelha do sisal (*Agave sisalana*), ao invés de *Aspergillus niger*, do qual era uma espécie críptica. Desde então, buscamos entender mecanismos de patogenicidade deste fungo através do sequenciamento do genoma de dois isolados provenientes de tecidos do caule infectado de *A. sisalana* e posterior análise do conteúdo gênico destes fungos. Os genomas sequenciados em Illumina HiSeq 2500 foram montados e anotados, possuindo aproximadamente 34,8 Mbp contendo 12.549 genes codificadores de proteínas e 32,2 Mbp contendo 11.809 genes codificadores de proteínas para CCMB663 e CCMB674, respectivamente. Os arquivos resultantes foram utilizados para a predição do secretoma, nos quais diferentes programas foram aplicados como filtros para predição das proteínas secretadas para o meio extracelular, e essas proteínas foram associadas à termos de ontologia gênica (GO). Nossos resultados mostram que há uma grande quantidade de termos GO que descrevem funções associadas à degradação de proteínas, lipídios, e principalmente, carboidratos, e que a abundância relativa destes termos é similar entre os diferentes isolados. A partir deste conjunto de proteínas preditas como secretadas, foram identificadas ainda possíveis proteínas efetoras, ou seja, proteínas que facilitam a invasão e colonização da planta infectada. Os resultados mostram que a maior parte das proteínas identificadas não tem similaridade com efetores conhecidos e, naqueles onde foi possível indentificar domínios conservados, há genes essenciais e de grande impotância para a virulência em plantas, os quais atuam tanto silenciando a resposta imune quanto a estimulando, podendo causar morte celular nos tecidos da planta. Foram analisadas ainda proteínas degradadoras de parede celular vegetal, secretadas e citosólicas. Há uma grande riqueza e quantidade de genes codificantes de proteínas degradadoras de parede celular, algumas especializadas na degradação de açúcares presentes em grandes quantidades nos tecidos do caule de sisal, além de proteínas de degradação que também agem como efetores. Com nossos resultados, ampliamos a compreensão do patossistema *Aspergillus welwitschiae* x *Agave sisalana* como também identificamos novos efetores de interesse na interação fungo-planta.



**PALAVRAS-CHAVE:** Fungo, Podridão Vermelha, Interação Planta-Patógeno.

**ABSTRACT (in English)**

In 2018, *Aspergillus welwitschiae* was described as the causing agent of the bole rot disease of sisal (*Agave sisalana*), rather than *Aspergillus niger* as previously thought, of which *A. welwitschiae* is a cryptic species. Since then, we have sought to understand the mechanisms of pathogenicity of this fungus by sequencing the genome of two isolates from *A. sisalana* infected stem tissues and subsequent analysis of the gene content of these fungi. The genomes sequenced in Illumina HiSeq 2500 were assembled and annotated, possessing approximately 34.8 Mbp containing 12,549 protein-coding genes and 32.2 Mbp containing 11,809 protein-coding genes for CCMB663 and CCMB674, respectively. The resulting files were used for the analysis of secretomes, in which different software were applied as filters to select the secreted proteins for the extracellular medium, and these proteins were associated with terms of Gene Ontology (GO). Our results show that there are a number of terms that describe functions associated with the degradation of proteins, lipids, and especially carbohydrates and that the relative abundance of these terms is similar between different isolates. From this set of proteins predicted as secreted, possible effector proteins were identified, that is, proteins that facilitate the invasion and colonization of the infected plant. Our results show that most of the proteins identified have no similarity with known effectors, and in those where it was possible to identify conserved domains, there are essential and very important genes for virulence in plants, which can act both to silence the immune response and to stimulate it, which can cause cell death in plant tissues. Secretory and cytosolic plant cell wall degrading enzymes were also analyzed. There is a great diversity of genes encoding cell wall degrading enzymes, some specialized in the degradation of sugars present in large quantities in the tissues of the sisal stem, as well as degradation proteins that also act as effectors. With our results, we have improved the understanding of the *Aspergillus welwitschiae* x *Agave sisalana* pathosystem, and also identified new effectors of interest in the fungus-plant interaction.

**KEYWORDS:** Fungus, Bole Rot, Plant-Pathogen Interaction.

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

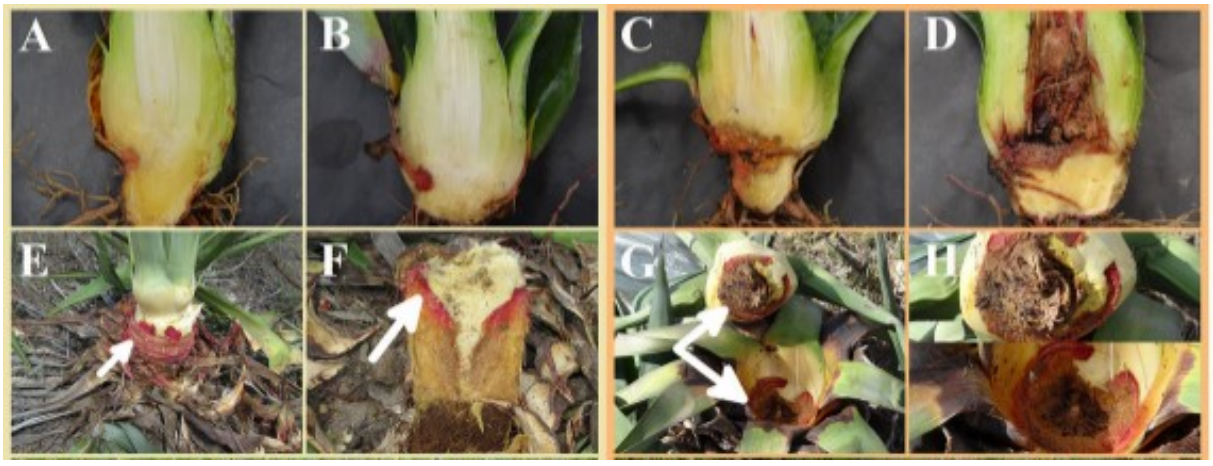
| | |
|---|---|
| AW663 | *Aspergillus welwitschiae* CCMB663 |
| AW674 | *Aspergillus welwitschiae* CCMB674 |
| AW139 | *Aspergillus welwitschiae* CBS 139.54 |
| bp | Base pairs |
| CAZyme | Carbohydrate-Active Enzyme |
| CBS 139.54 | *Aspergillus welwitschiae* CBS 139.54 |
| CCMB663 | *Aspergillus welwitschiae* CCMB663 |
| CCMB674 | *Aspergillus welwitschiae* CCMB674 |
| CWDE | Cell Wall-Degrading Enzymes |
| EC | Enzyme Commision |
| ETI | Effector-Triggered Immunity |
| GO | Gene Ontology |
| HR | Hypersensitive Reaction |
| HST | Host-Specific Toxin |
| JGI | Joint Genome Institute |
| NB-LRR | Nucleotide-Binding Leucine-Rich Repeat |
| PAMP | Pathogen-Associated Molecular Pattern |
| PDA | Potato-Dextrose Agar |
| PTI | PAMP-Triggered Immunity |
| v | version |

## TABLE OF CONTENTS

## 1 INTRODUCTION

*Aspergillus welwitschiae* is a filamentous fungal species in the phylum Ascomycota (ESSER, 2001). Its spores are ubiquitous in soil, air, and domestic environments (VARGA et al., 2014). In nature, *A. welwitschiae* acts mainly as a saprotroph, feeding on dead plant matter. The misidentification of *Aspergillus welwitschiae* for its closest species, *Aspergillus niger*, is very common since the two species are morphologically indistinguishable, except with the employment of molecular techniques (GHERBAWY et al., 2015). This erroneous identification led *Aspergillus niger* to be considered on different occasions as the causative agent of the bole rot of sisal, a plant of the *Agave* genus (MHAIKI et al., 2013; SANTOS et al., 2014). Both studies use molecular identification through genomic regions that are not divergent enough to differentiate species in the Aspergillus genus (the ITS region and the beta-tubulin gene), which is differentiated with the calmodulin gene. Nonetheless, Duarte et al. described *A. welwitschiae* as the true causative agent of sisal's bole rot disease (DUARTE et al., 2018), using the adequate methods of identification. In this disease, the fungus invades the plant through wounds in the leaf basis, caused by the leaf's extraction, valued for its natural fibers. *A. welwitschiae* has also been found in dried fruits and nuts, some isolates producing secondary metabolites that pose a treat to human health, such as fumonisins and ochratoxins (MASSI et al., 2016; SUSCA et al., 2014). Those metabolites may be part of the fungus strategy to infect plants and feed on its components (BENNETT; KLICH, 2003).

**Figure 1:** Progression of the bole rot disease in Agave sisalana. Arrows indicate the colonization of *Aspergillus welwitschiae* in the bole tissue. Adaped from Duarte et al., 2018.




As described by Duarte, the bole rot of sisal compromises a very important cultivation in one of the poorest regions of Brazil, in the Caatinga. This biome is

characterized by the scarcity of water supplies, low rain incidence and high temperatures (ALVES; ARAÚJO; NASCIMENTO, 2009). *Agave sisalana* is particularly suited for this climate, considering its origins in the Yucatán peninsula, in Mexico (DEBNATH et al., 2010). Because of the harsh conditions in the Caatinga biome region, sisal represents one of the few available cultivation options in this area and employes people in family-based farmlands. In some cases, rejected leaf material from the fiber extraction process is served to sheep and goats farming in these same farmlands (BOTURA et al., 2011; SILVEIRA et al., 2012).

According to the Food and Agriculture Organization of the United Nations (FAOSTAT), the production of sisal is currently in decline in Brazil, however, the country remains the leading producer of hard natural fibers derived from the plant, followed by Tanzania (KIMARO; MSANYA, 1994). This decline is partly due to the losses in yield because of the bole rot disease. In fact fungal diseases are a great threat to crops worldwide (LI et al., 2015). Besides its clear social and economic relevance, pathosystems like the bole rot involve many mechanisms and enzymes that might be valuable to the biotechnological industry. *Aspergillus niger*, the closest species to *A. welwitschiae* and most common species of the genus, is largely used in the industry as a cell factory (BRAAKSMA et al., 2010), that is, a living organism used for the production of complex molecules. For example, in 1918 *A. niger* has been described as a producer of citric acid when grown in high concentrations of sugar, and has since led the industrial production of this compound (CAIRNS; NAI; MEYER, 2018). As the closest species, it is very likely that *Aspergillus welwitschiae* is capable of the same metabolic processes that made *A. niger* relevant in the biotechnological industry.

*Agave sisalana* is a very tolerant crop, resisting to the dry conditions of Caatinga, and also known for its ability to absorb heavy metals (DOS SANTOS et al., 2011), production of antimicrobial (SANTOS et al., 2009) and anti-nematode compounds (SILVEIRA et al., 2012). We hypothesize that to invade and colonize such host, *Aspergillus welwitschiae* possesses a strong arsenal of carbohydrate-active enzymes, to degrade the hard fibers of sisal, and also effector proteins, to manipulate the plant immune responses among other strategies that might help this fungus to induce disease in sisal.

# 2 LITERATURE REVIEW

## 2.1 Plant-Pathogen Interaction

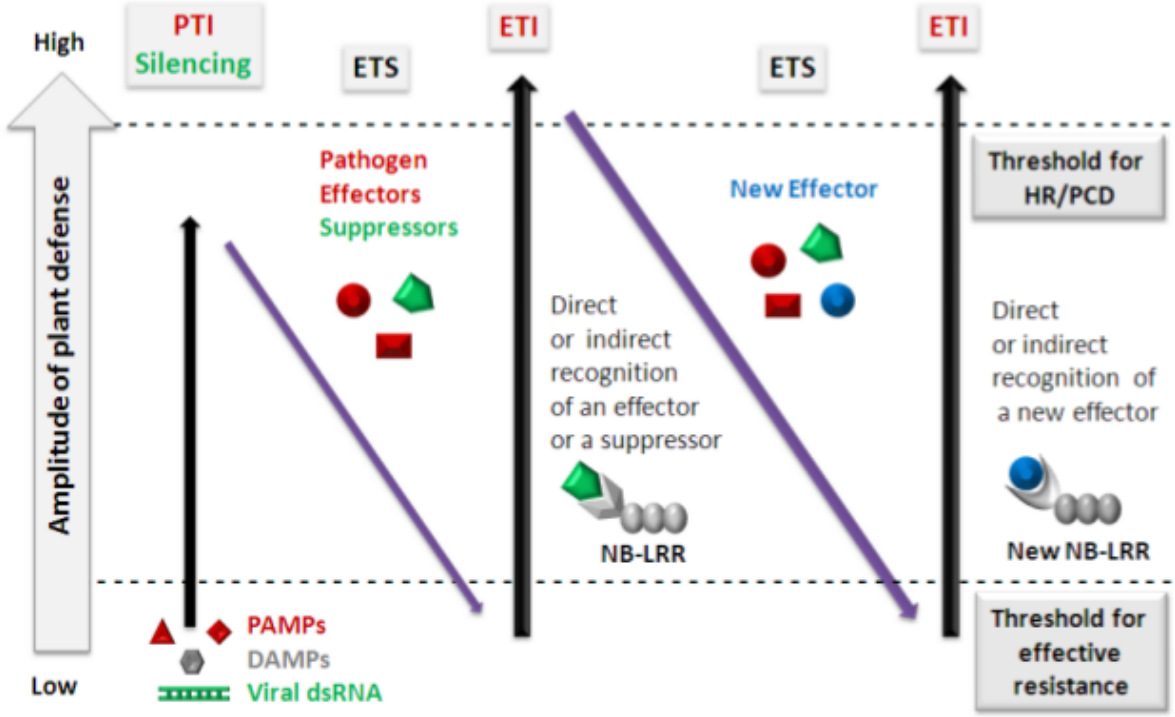
Plants face constant biotic stress caused by pathogenic microorganisms (DODDS; RATHJEN, 2010). Those pathogens, such as bacteria (MANSFIELD et al., 2012), viruses (PALLAS; GARCÍA, 2011) and fungi (ROSE et al., 2014) pose a great threat to crops across the globe, causing great economic loss. Fungi are one of the most damaging groups of phytopathogens, affecting many crops of economic interest like rice, wheat and corn (DEAN et al., 2012).

Throughout the co-evolution of plants and pathogens, plants developed a complex multilayered defense system (KOECK; HARDHAM; DODDS, 2013), including primary physical barriers, such as the cell wall and cuticle (BENOIT et al., 2015), and also two layers of pathogen-induced response. The first layer, known as PAMPs-Triggered Immunity (PTI – “PAMPs” meaning *pathogen-associated molecular patterns*) is a broader response activated by molecules present in pathogens, such as flagellin in bacteria and chitin (a polymer that constitutes the fungal cell wall) in fungi (SONAH; DESHMUKH; BÉLANGER, 2016). PTI unleashes responses such as reorganization of the cytoskeleton, fortification of the cell wall, synthesis of phytoalexins, oxidative burst and the Hypersensitive Response (HR), which consists of local programmed cell death on the infection recognition site, and is meant to limit the spread of the pathogen (ASHWIN et al., 2018). However, fungi have means of avoiding this line of defense, which leads to the second, most specific line of defense, called Effector-Triggered Immunity (ETI) (ZVEREVA; POOGGIN, 2012). Effectors are proteins that enable pathogens to manipulate the host cellular environment, by targeting the natural defenses and creating ideal conditions for growth and reproduction. To fight against this strategy, plants possess R-genes (R meaning *resistance*) that counter-target effector proteins targeting plant defenses (WANG et al., 2014). Figure 2 shows a scheme of how the plant response is modulated in the presence of a pathogen.

To invade a plant host and overcome its innate defenses, the fungi possess very diverse strategies that vary according to that fungal niche. The phytopathogenic fungi are commonly divided into biotrophic, hemibiotrophic, and necrotrophic (DOEHLEMANN et al., 2007). Biotrophic fungi require living tissues to obtain nutrients

from and limit the quantity of secreted effectors and lytic enzymes to keep tissues alive (RODRIGUEZ-MORENO et al., 2018). Hemibiotrophic fungi are similar to biotrophs, however, they kill the host's cell in latter stages to complete their life cycle (KOECK; HARDHAM; DODDS, 2013). Necrotrophs can be subdivided into broad-host-range (GUYON et al., 2014) and host-specific species, which secrete host-specific toxins (HSTs) (MORETTI; SUSCA, 2017) without which they lose pathogenicity to that specific host. *Aspergillus welwitschiae* is a broad-host-range necrotrophic fungus,

**Figure 2:** Innate immunity in plants. Pathogen- or Damage-associated patterns trigger PAMPs-Triggered Immunity (PTI), that is then silenced by effector proteins, causing effector-triggered susceptibility (ETS). Effector recognition by the plant host's nucleotide binding site-leucine rich repeat (NB-LRR) results in Effector-Triggered Immunity, which might be silenced by new effectors, leading to successful colonization, or in case of recognition, lead to plant resistance or hypersensitive reaction and programmed cell death. Adapted from Zvereva and Pooggin, 2012.




being detected in dried fruits, seeds, and sisal, the study model of this work.

Briefly, the strategy of necrotrophic fungi to kill the host cell relies heavily upon an oxidative burst, the secretion of low molecular weight peptides and cell wall degrading enzymes, which will be described in detail in the next topics. All these processes together collapse the host cells and provide nutrients for the fungi to grow (LORANG, 2018).

## 2.2 Fungal Genomics

The study of genomes (genomics) originated with the advent of DNA sequencing technologies. For fungi, the first genome to be sequenced was that of the yeast *Saccharomyces cerevisiae* in 1996 (GOFFEAU et al., 1996) still using the Sanger method, before the Human Genome Project, in which the draft results were first published in 2001 (VENTER et al., 2001), followed by many other species and initiatives, such as the 1000 fungal genomes initiative, developed by the Joint Genome Institute (JGI) (GRIGORIEV et al., 2014), which was made possible by the modern high-throughput sequencing technologies, lead by Illumina. Fungal genomics studies shed light into different perspectives of fungal biology, such as cryptic species, which could not be distinguished by morphology or metabolic tests (AKAGI et al., 2009), and also mating (GIOTI et al., 2012), which still is an incognita in fungal biology, considering that many fungal groups do not have a sexual phase described, as the genus Aspergillus section Nigri (ALVAREZ-PEREZ et al., 2010). In summary, fungi have an asexual reproduction stage, in which they produce genetic variability, yet little, by parasexuality processes and also a sexual stage, in which different mating types create a new combination of genetic material by karyogamy, that is, the fusion of somatic cells forming a zygote.

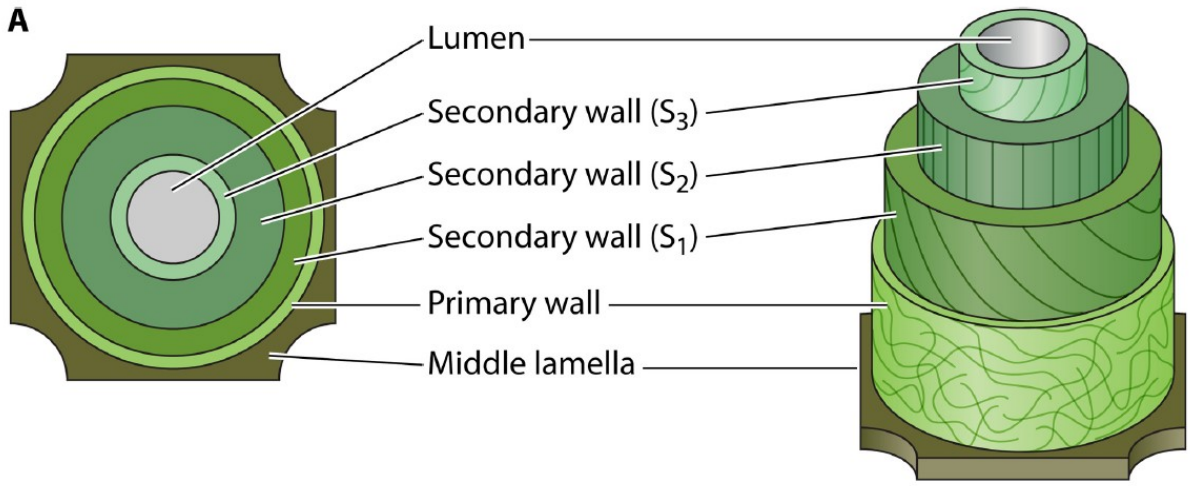
Besides phylogenetics and genotyping studies, fungal genomics studies are important in molecular biology for the description of new gene functions that might have value in the biotechnology industry and enzymatic processes that help us to understand crop susceptibility to fungal diseases. In the Aspergillus genus, *Aspergillus niger* is extensively studied as a “cell factory” for agriculture industrial waste treatment. Studies characterize the production of enzymes such as proteases (CHELLAPANDI, 2010; PARANTHAMAN; ALAGUSUNDARAM; INDHUMATHI, 2009), pectin methylesterase (JIANG et al., 2013), pectinases (EL-SHEEKH et al., 2009), and endo-β-1,4 mannanase (NAGANAGOUDA; SALIMATH; MULIMANI, 2009) from food waste, and also its use as a biological control against aflatoxin contamination (KING et al., 2011), and removal of polyphenolic compounds from the coffee industry (MACHADO et al., 2012), to cite only a few examples.

## 2.3 Polysaccharide degradation in fungi

One of the most valued molecules of fungal origin are the carbohydrate-degrading enzymes due to their direct application in biofuels, paper and food industries

(VAN DEN BRINK; DE VRIES, 2011). For the fungi, these enzymes are used to degrade very complex molecules, such as lignin (POLLEGIONI; TONIN; ROSINI, 2015), which is full of aromatic rings making its degradation considerably difficult. Furthermore, these enzymes are also used to degrade hydrocarbons, such as in *Yarrowia lypolytica,* that possess the potential to mitigate oil spills (ZINJARDE; PANT, 2000). Therefore, fungi represent an alternative to the degradation of compounds that otherwise might be very complex and expensive to perform.

**Figure 3:** Representation of the structure of plant the cell wall, including the middle lamella, primary and secondaries cell walls and the lumen. Adapted from Rytioja et al. (2014).




For the fungi itself, carbon-degrading enzymes are essential, especially in saprophytic fungi (as our model of study, *Aspergillus welwitschiae*) that rely on a specific group of carbon-degrading enzymes, the cell wall-degrading enzymes (CWDE)(STRUCK, 2006). Besides the previously mentioned lignin, which is part of plants secondary cell wall (POLLEGIONI; TONIN; ROSINI, 2015), the cell wall is composed of different layers of carbohydrates in many different combinations and arrangements (Figure 3). The innermost layer of the primary cell wall of plants is composed of cellulose, and adhering polysaccharides, such as xyloglucans, glucuronoarabinoxylans or glucomannans (PETTOLINO et al., 2012). The following layer consists of a gel-like pectin structure, mixed with other polysaccharides (JIANG et al., 2013). The final layer is composed of structural proteins (KEEGSTRA, 2010). Nonetheless, other plant structures, such as the cuticle (which is also a physical barrier against pathogens) is also made of lipids, in this case, cutin (ZHAO et al., 2013). The middle lamella, which can be considered the outermost layer of the cell wall, is composed of calcium and magnesium pectates (JIANG et al., 2013). CWDE is a large

group that includes proteins for the degradation of all supra mentioned polysaccharides and others that might make part of the cell wall in different tissues or conditions (RODRIGUEZ-MORENO et al., 2018). Those enzymes are specialized in the breakage of the bonds between those polysaccharides, transforming long chains in oligomers that the fungi can absorb (KAVANAGH, 2011). Thus, some carbohydrate degradation enzymes (in general) have to be secreted to reach the plant host cell, while others are cytosolic, working on the final steps of degradation from inside the fungal cell.

### 2.4 Effectors

Fungal “effector protein” is a very broad concept, meaning proteins that are secreted to the host environment, including the apoplast or cytosol (in plants) that facilitate the fungi entrance and colonization of that host (KOECK; HARDHAM; DODDS, 2013). In that sense, cell wall degrading enzymes (CWDE) described in the previous topic can also be considered effectors, when they are not cytosolic (EXPERT et al., 2018; FU et al., 2015; YAKOBY et al., 2007). The criteria to be considered an effector protein include proteins with a signal peptide for secretion, no transmembrane domains, no similarity with other obvious protein domains, fairly small size and (in some cases) species-specificity (SONAH; DESHMUKH; BÉLANGER, 2016), which includes the host-specific toxins (HST) previously mentioned. Therefore, studies to identify candidate fungal effectors are based on looking for uncharacterized proteins, meaning that those proteins that have no identifiable protein domains or annotated gene functions in databases like KEGG or Gene Ontology.

In the literature, plant effectors have been described in many plant pathogenic fungi. Some examples include an study that characterize the set of candidate secreted fungal effectors in the fungus *Moniliophthora perniciosa*, causative agent of the witches broom disease of cacao (BARBOSA et al., 2018), and another similar study for the phytopathogen *Sclerotinia sclerotiorum*, which causes the white mold in more than 400 species (GUYON et al., 2014), and also another study, which described the effectorome of three species of the genus Trichoderma (GUZMÁN-GUZMÁN et al., 2017). As a common point, these studies found different cell wall-degrading enzymes, but also elicitors (proteins that bind to membrane receptors initiating a signaling cascade).

# 3 OBJECTIVES

## 3.1 General Objective

Our objective was to compare the genomes of three isolates of *Aspergillus welwitschiae*, two of which (isolates CCMB663 and CCMB674) were isolated from sisal (*Agave sisalana*) in the *Sisaleira* (sisal-producing) region of the state of Bahia, the main-producer region of hard natural fibers globally, and the other from the ex-culture type specimen (isolate CBS 139.54) provenient of *Welwitschia mirabilis* in the Namib Desert in Namibia.

## 3.2 Specific Objetives

- To assemble and annotate two new nuclear and mitochondrial genomes of *Aspergillus welwitschiae* (isolates CCMB663 and CCMB674), the causative agent of the red rot of sisal.
- To compare and analyze the predicted secretome, effectorome and cell wall-degrading enzymes content of *Aspergillus welwitschiae* isolates CCMB663, CCMB674 and CBS 139.54 to better understand mechanisms of pathogenicity of *Aspergillus welwitschiae* in sisal plants.

# 4 MATERIALS AND METHODS

## 4.1 Biological Material

This study comprises three isolates of *Aspergillus welwitschiae*: CCMB663 and CCMB674, which were directly isolated from field-collected *Agave sisalana* individuals in different locations (Table 1) (DUARTE et al., 2018), and also CBS 139.54, originally isolated from *Welwitschia mirabilis* in the Namib Desert, available at the JGI Genome Portal (VESTH et al., 2018).

**Table 1:** *Aspergillus welwitschiae* isolates specifications

| Isolate | Haplotype | GenBank | Location |
|---|---|---|---|
| CCMB674 | Hap6 | MG322278 | São Domingos, Brazil |
| CCMB663 | Hap1 | MG322270 | Conceição do Coité, Brazil |
| CBS 139.54 | Hap1 | GCA_003344945.1 | Namib Desert, Namibia |

The mycelium of isolates CCMB663 and CCMB674 was grown on potato dextrose agar medium (PDA – Sigma-Aldrich, Missouri, USA) and incubated at 25 °C for 5 days or after covering the surface of a 9-mm diam Petri dish, and after that, it was scrapped for genomic DNA extraction, which was performed with FastDNA™ for Soil kit (MP Biomedicals – California, USA). Genomic DNA quality and quantity were assessed by agarose gel electrophoresis and fluorometric analysis, respectively. A 450 bp library was prepared with NEBNext Fast DNA Fragmentation and Library Preparation Kit (New England Biolabs, Nebraska, USA) following the manufacturer's instructions. Library quality was evaluated with a 2100 Bioanalyzer (Agilent – California, USA), and whole-genome sequencing was performed on a HiSeq 2500 platform with pair-end strategy and estimated fragment size of 450 bp (Illumina – California, USA) in Georg August University Göttingen.

## 4.2 Nuclear Genome Assembly

### 4.2.1 Genome Assembly

The quality of 8,400,440 and 17,955,623 raw paired-end reads, for CCMB663 and CCMB674 respectively, were assessed using FastQC v0.11.5 (ANDREWS, 2010), and low-quality bases (Phred score <20) were trimmed and overlapping sequences were collapsed using AdapterRemoval v2 with default parameters (LINDGREEN,

2012). FASTQ output files were partitioned with the tool GCSplit v1.2 (MIRANDA et al., 2018) into 4 sets of reads with similar GC content, and those partitioned files were assembled with SPAdes v3.11.1 (BANKEVICH et al., 2012) with the set of k-mers 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87 and the --careful flag.

### 4.2.1 Genome Quality Control and Annotation

Resulting assemblies were annotated with MAKER2 v2.31.9 (YANDELL; HOLT, 2011), with support from ab initio predictors SNAP v2006-07-28 and Augustus gene prediction program (STANKE et al., 2004). To improve gene prediction, a protein set from *Aspergillus niger*, the phylogenetically closest species to *Aspergillus welwitschiae* (PALUMBO; O'KEEFFE, 2015) was also used. Assembly QC was analyzed with Perl script *scaffold_stats.pl* (THORPE et al., 2018) and BUSCO v3, using the Ascomycota lineage dataset, available at the tool's website (SIMÃO et al., 2015). Genome sinteny was accessed with Mauve v20150226 (DARLING et al., 2004).

## 4.3 Mitochondrial Genome Assembly

Mitochondrial Genome was assembled separately from the nuclear genome for both CCMB663 and CCMB674. We accessed different techniques to recover mitochondrial contigs (ideally, a single contig) from our raw reads. Both approaches (described in detail in the following topics) started with the cleaning of sequencing reads with BBDuk, part of the BBTools v38.30 software package (BUSHNELL, 2018) for the removal of reads below Phred score 20, parameters including both the left and right ends of the sequence, with the paired reads being processed together. Clean reads were used for posterior analysis.

### 4.3.1 Mapping-based Genome Assembly

Inspired by methods that rely on reference genomes to assembly mitochondria (DIERCKXSENS; MARDULYN; SMITS, 2017), we developed a workflow to recover and assembly *de novo* mitochondrial reads. We used the only publicly deposited *Aspergillus welwitschiae* mitochondrial genome (isolate CBS 139.54), which is available on JGI (Joint Genome Institution, https://jgi.doe.gov/). This reference genome was formatted and used for alignment with Bowtie2 v2.3.4.3 (LANGMEAD et al., 2019) against the paired-end sequencing reads. Aligned reads were output to a .FASTQ file

and *de novo* assembled with SPAdes v3.12 with default parameters (BANKEVICH et al., 2012) trying to reconstitute the whole mitochondrial genome.

### 4.3.2 Direct Genome Assembly

This methodology was modified from that of De Vries et al. (DE VRIES; GRIGORIEV; TSANG, 2011), and was based on the assembly and recovery of the mitochondrial genome through comparison with known mitochondrial genomes. Clean reads were directly assembled *de novo* with SPAdes v3.12, and mitochondrial contigs were identified with BLASTN, part of the BLAST+ v2.8.1 tool suite (CAMACHO et al., 2009), using Mitochondrial RefSeq vDecember-2018 as database (available at ftp://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/). Relevant contigs, meaning contigs presenting hits with more than 90% similarity with mitochondrial genomes and length around the expected genome size, were selected as part of the mitochondrial genome.

### 4.3.3 Genome Quality Control and Annotation

Assembly statistics were accessed with *scaffold_stats.pl.* Results with expected characteristics (genome size and contig number) were selected for genome annotation. To perform these analyses, we used the MITOS2 web server v2017-09-02 (BERNT et al., 2013) for general genome information, RNAweasel (BECK; LANG, 2009) for the correct presence of tRNA genes and intronic regions, and MFannot (BECK; LANG, 2010) for ORF presence. The resulting GFF file was used to construct a graphic representation of genomes, performed in Geneious v11.0.5 (KEARSE et al., 2012) and Microsoft Excel. A guiding phylogenetic tree (Maximum Likelihood) was constructed on MEGA v7.0.26 (KUMAR et al., 2008) using the rrnS gene sequence, with the Jukes-Cantor method and 500 Bootstrap replications..

## 4.4 Secretome analysis

Our approach to identify secreted proteins was modified from Cuesta-Astroz et al. (2017). All annotated proteins were submitted to SignalP v4.1 (PETERSEN et al., 2011), which identifies classical secretory proteins. Possible mitochondrial proteins were removed by using TargetP v1.1 (EMANUELSSON et al., 2007) and sequences containing transmembrane helices were discarded, using TMHMM v2.0, a web application based on Hidden Markov Model method (KROGH et al., 2001), with default parameters for the three applications. Putative secreted proteins were mapped to

Gene Ontology (GO) terms and annotated using GO Feat (ARAUJO et al., 2018). GO Feat annotation data was imported, and a Venn Diagram tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to describe the distribution of those GO terms in our three isolates. We used this kind of visualization to guide our following analyses. The relative abundance of shared GO terms was generated with python script *count.py* and Microsoft Excel 2016, and heatmaps were produced on R v3.5 (R CORE TEAM, 2013). This approach was repeated for GO terms shared by all three isolates, and also the ones shared only by CCMB663 and CCMB674, or between each of these isolates and CBS 139.54. Exclusive-occurring GO terms were mapped back to their associated secreted proteins manually, and, subsequently, those proteins were compared against the non-redundant database with BLASTP, part of the BLAST+ v2.8.1 tools suite (CAMACHO et al., 2009) This annotation was manually curated and associated to their respective Enzyme Commission (EC) number, or classified as “Unavailable” when this number was not found or when the protein-coding gene does not code for an enzymatic protein. For both color matrixes and tables, redundant terms were collapsed (that is, summed to one main category, now described as “concise”) and terms referring to cellular components were discarded. As we applied different filters to assure that all proteins are addressed to the extracellular space, a cellular location description is not necessary and might be misleading.

### 4.5 Effector Proteins Analysis

For our effector proteins analysis, it is worth noting that we considered effectors as a subset of the secretome since those proteins act directly on the host, outside of the fungal cell. Our analysis started by submitting all secreted proteins in each predicted isolate secretome to EffectorP v2.0 (SPERSCHNEIDER et al., 2018). Considering this web application is not based on protein similarity, we submitted predicted effectors to PHI-base v4.7 (URBAN et al., 2017) to get more information on those proteins. After that, we submitted effectors with relevant hits to IUPred2 v2018-Dec-17 (MÉSZÁROS; ERDÖS; DOSZTÁNYI, 2018) and ANCHOR2 v2018-Dec-17 as described by Kuppireddy (2017) to describe Pfam domains (included on IUPred2A prediction), and regions of intrinsic disorder, and also possible binding sites of target proteins.

### 4.6 Carbohydrate-Active Enzymes Analysis

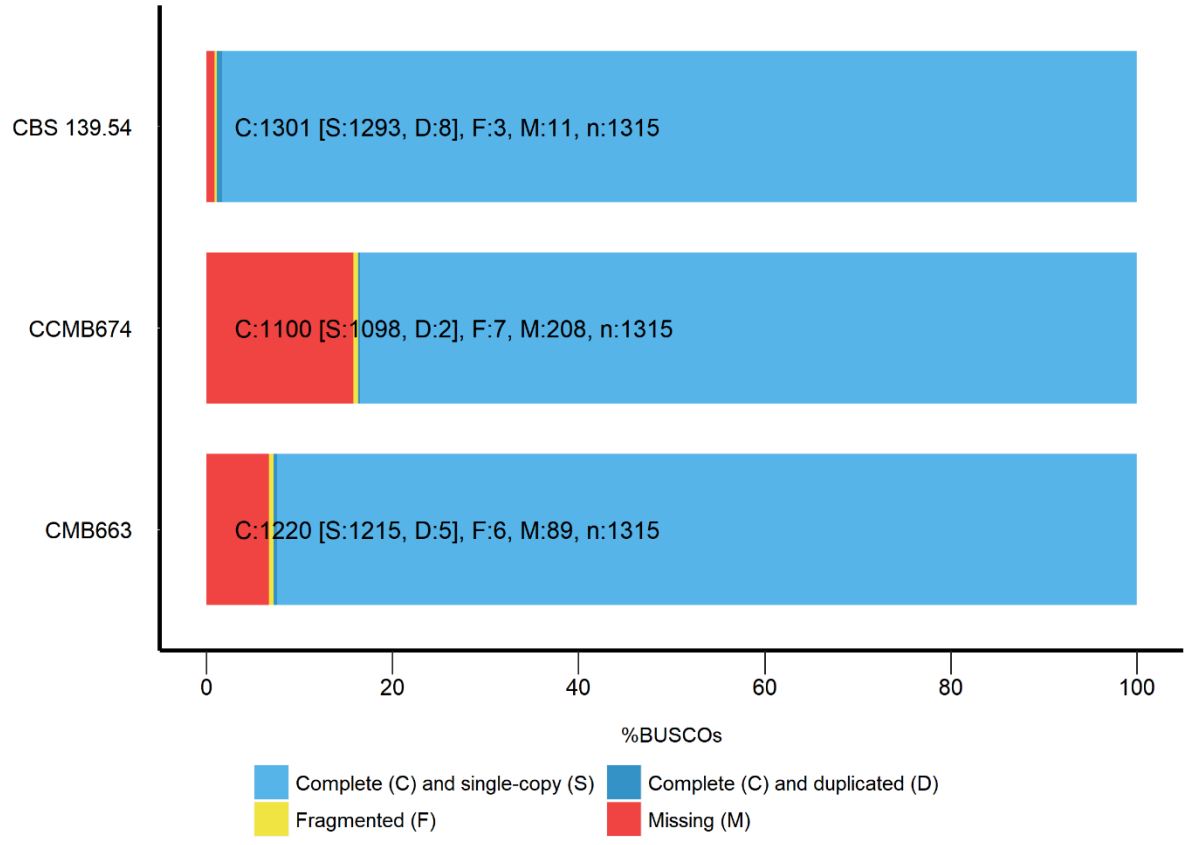
Carbohydrate-active enzymes were identified with dbCAN v2 (ZHANG et al., 2018) based in protein-coding genes predicted with MAKER2 v2.31.9 (YANDELL; HOLT, 2011) for CCMB663 and CCMB674. For the analysis of CBS 139.54, we obtained the protein file to submit for dbCAN v2 from JGI. For protein identification, dbCAN v2 applies the HMMER search against the dbCAN HMM (hidden Markov model) database v7, followed by a DIAMOND search against the CAZy pre-annotated CAZyme sequence database v7-31-2018, and a Hotpep search against the conserved CAZyme short peptide database. The results were joined, and redundant enzymes were removed from the output. Lastly, the frequency of each carbohydrate-active enzyme was calculated for comparison among the three *A. welwitschiae* isolates.

# 5 RESULTS

## 5.1 Nuclear Genome Assembly Analysis

The nuclear genome for our isolates CCMB663 and CCMB674 was assembled and annotated as described, and then quality control was accessed. The quantitative measure of genomic data completeness in terms of expected gene content revealed that one of our *A. welwitschiae* isolates, CCMB663, is more similar when compared to the reference strain CBS 139.54 than our other isolate, CCMB674. Most of the predicted genes are complete and single-copy in the three *A. welwitschiae* isolates evaluated, although the genetic content in the reference isolate CBS 139.54 is greater than our two isolates. The number of missing genes is increased in the *A. welwitschiae* CCMB674, in which this amount is more than the double than detected in the other isolates. The number of fragmented and duplicated genes is similar in all isolates (Figure and Table 1). Using our workflow, we identified 12,549 protein-coding genes in the 34.8 Mb genome of the CCMB663, and 11,809 protein-coding genes in the 32.2 Mbp genome for CCMB674, while the genome of the reference isolate (CBS139.54) is

**Figure 4:** BUSCO quality assessment results for *Aspergillus welwitschiae* CCMB663, CCMB674 and CBS139.54. Colors and abbreviations are described at the bottom of the figure.



37.5 Mbp size with 13,687 protein-coding genes. Genome sintenyanalysis with Mauve (Figure 5) show that genomes are similar in concent but not syntenic. That is, the genomic content is similar but the organization is different.

**Figure 5:** Nuclear genome alignment with Mauve. Color blocks and connecting lines represent conserved regions in the three genomes. Rows show CCMB663, CCMB674 and CBS 139.54.

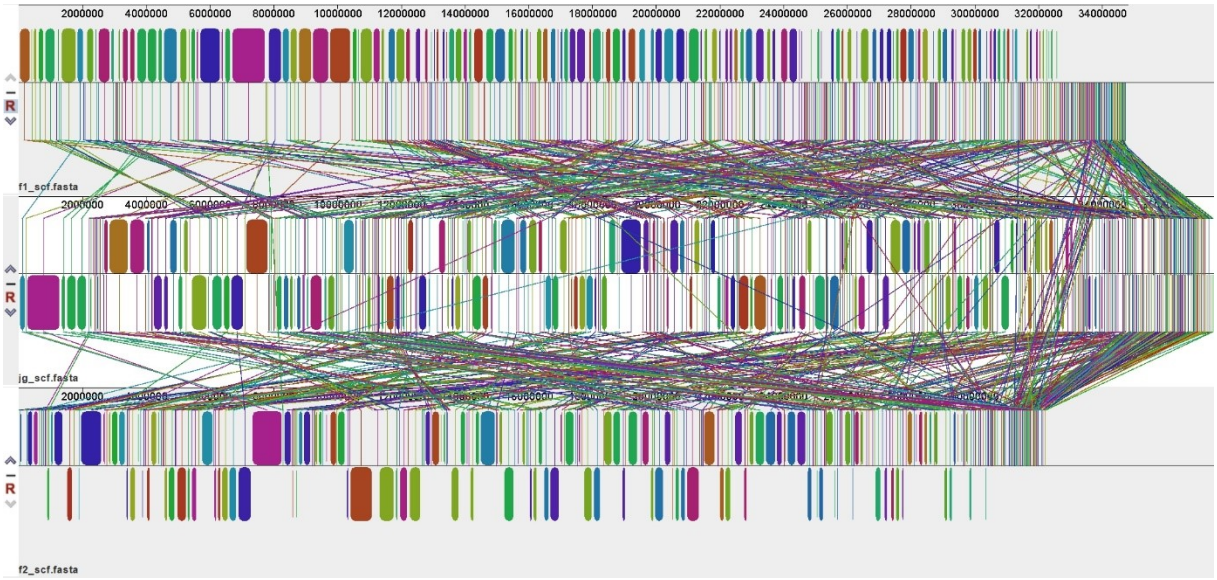



**Table 2:** Nuclear Genome assembly statistics.

| Parameter | CCMB663 | CCMB674 | CBS 139.54 |
|---|---|---|---|
| **Contigs >0** | 584 | 192 | 396 |
| **Contigs >1000** | 194 | 16 | 396 |
| **Total Span** | 34,819,369 | 32,282,741 | 37,511,876 |
| **Min. Contig** | 72 | 506 | 1010 |
| **Mean contigs** | 59,662 | 168,139 | 94,726 |
| **N50** | 579,553 | 675,617 | 757,052 |
| **L50** | 18 | 16 | 17 |
| **% GC** | 0.495 | 0.493 | 0.497 |

### 5.2 Mitochondrial Genome Assembly Analysis

In this stage, two different methods were employed for mitochondrial genome reconstruction. Between the mapping based and direct assembly, the latter presented the best results. Direct assembly recovered a single mitochondrial contig, for both isolates CCMB663 and CCMB674, and this contig has a greater length than the sum of the three contigs obtained for both isolates with the mapping-based approach. In this strategy, three contigs were recovered using a reference for each isolate. The smaller contig in CCMB663 and CCMB674 is similar to that of CBS 139.54, while the other two contigs are similar to the 25410 bp contig of the reference genome. In all assemblies, a single contig represents more than 50% of the total length (L50 value) and our best result, with the direct assembly, is better than the assembly provided for

isolate CBS 139.54, which contains two contigs. Mitochondrial genome size is similar between those three genomes, with a short difference that might be due to the gap between the two contigs of CBS 139.54. Assembly values are seen in Table 3.

**Figure 6:** Genomic organization of essential mitochondrial genes and homing endonucleases in CCMB663, CCMB674, CBS 139.54 and A. niger N909. "*" indicates LAGLIDADG, "**" indicates GIY-YIG.

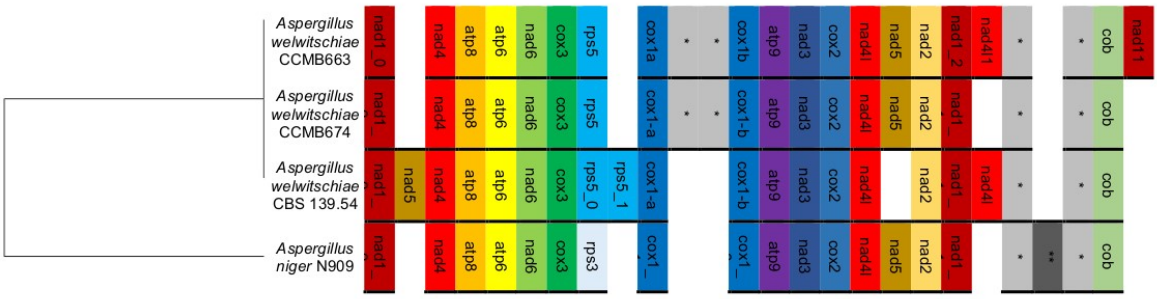



**Table 3:** Mitochondrial Genome assembly statistics.

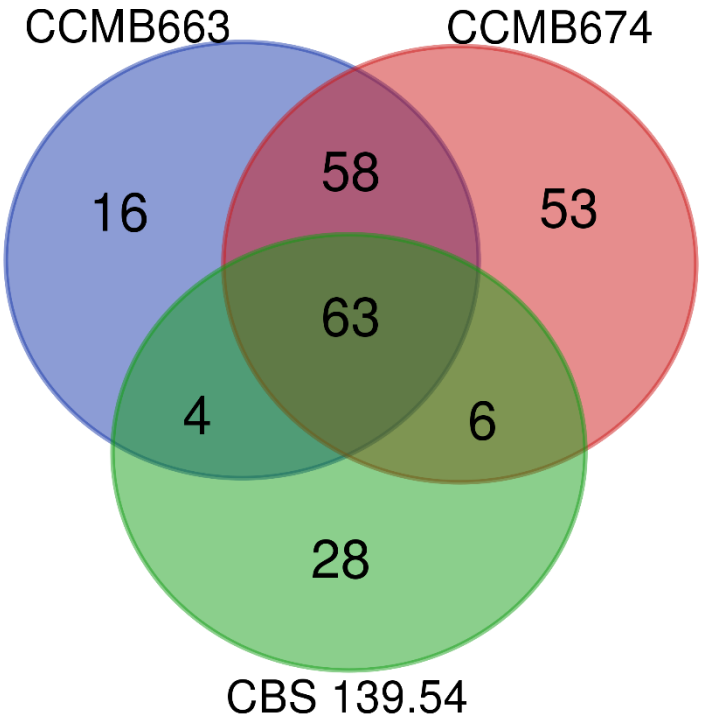
| | Mapping-based | | Direct assembly | | Reference |
|---|---|---|---|---|---|
| **Parameter** | **CCMB663** | **CCMB674** | **CCMB663** | **CCMB674** | **CBS 139.54** |
| **Total Contigs** | 3 | 3 | 1 | 1 | 2 |
| **Total Span** | 29675 | 29654 | 31472 | 31490 | 30950 |
| **Min. Contig** | 5583 | 5560 | 31472 | 31490 | 5540 |
| **Mean contigs** | 9891 | 9884 | 31472 | 31490 | 15475 |
| **N50** | 17384 | 17386 | 31472 | 31490 | 25410 |
| **L50** | 1 | 1 | 1 | 1 | 1 |
| **% GC** | 0.269 | 0.269 | 0.265 | 0.264 | 0.269 |

Overall, annotation of mitochondrial genomes described 23 protein-coding genes and 27 tRNA regions on CCMB663, while CCMB674 and CBS 139.54 both contain 21 protein-coding genes and the same amount of tRNA regions. The two different genes in CCMB663 are duplications in nad1 and nad4L. The phylogenetic tree based on rps3 only highlights the differences between species, *Aspergillus welwitschiae* and *Aspergillus niger*, since the protein sequences in the three isolates are identical. *A. niger* also contains a different homing endonuclease, absent in the other fungi, GIY-YIG.

### 5.3 Secretome Analysis

In this step, we evaluated proteins that could be secreted to the extracellular space and possibly be in contact with the host environment, having some effect on the invasion and pathogenesis process. Initially, we use Gene Ontology terms instead of the proteins' sequences, which were later recovered and analyzed for exclusive occurring terms. We identified 461 secreted proteins in CCMB663, 612 proteins in CCMB674 and 429 in CBS 139.54, including 226 unique GO terms, associated with 189 (40%), 274 (44%), and 168 (39%) of those putative secreted proteins in CCMB663, CCMB674 and CBS 139.54, respectively. The complete list of those GO terms can be found in Table S1. A total of 63 of those terms are shared by all of those isolates while 68 are shared by only two out of the three isolates, mostly involving CCMB663 and CCMB674, isolated from *Agave sisalana*.

**Figure 7:** Venn diagram of the distribution of 226 unique Gene Ontology terms in *Aspergillus welwitschiae* CCMB663, CCMB674 and CBS 139.54.




There are also 16, 53 and 28 exclusively occurring terms for CCMB663, CCMB674 and CBS 139.54, respectively. The GO terms descriptions allow us to see that most of these terms are related to the degradation of carbohydrates, proteins, and lipids, even in the exclusive terms. These terms are associated with the fungal feeding strategy and are related to the decomposition of dead matter. The relative abundance

of those terms is very similar between the three isolates, which highlights the similarity in the secretome content in those *Aspergillus welwitschiae* isolates.

Distribution of GO terms is shown in the Venn diagram (Figure 7), the relative abundance of terms present in all isolates, or only two isolates can be verified in the heatmaps (Figure 8 and 9), and exclusive occurring terms and their associated proteins can be found in Table 4.

**Figure 8:** Heatmap with Gene Ontology terms (rows) shared by CCMB663, CCMB674, and CBS 139.54 (columns). Darker colors indicate higher abundance.

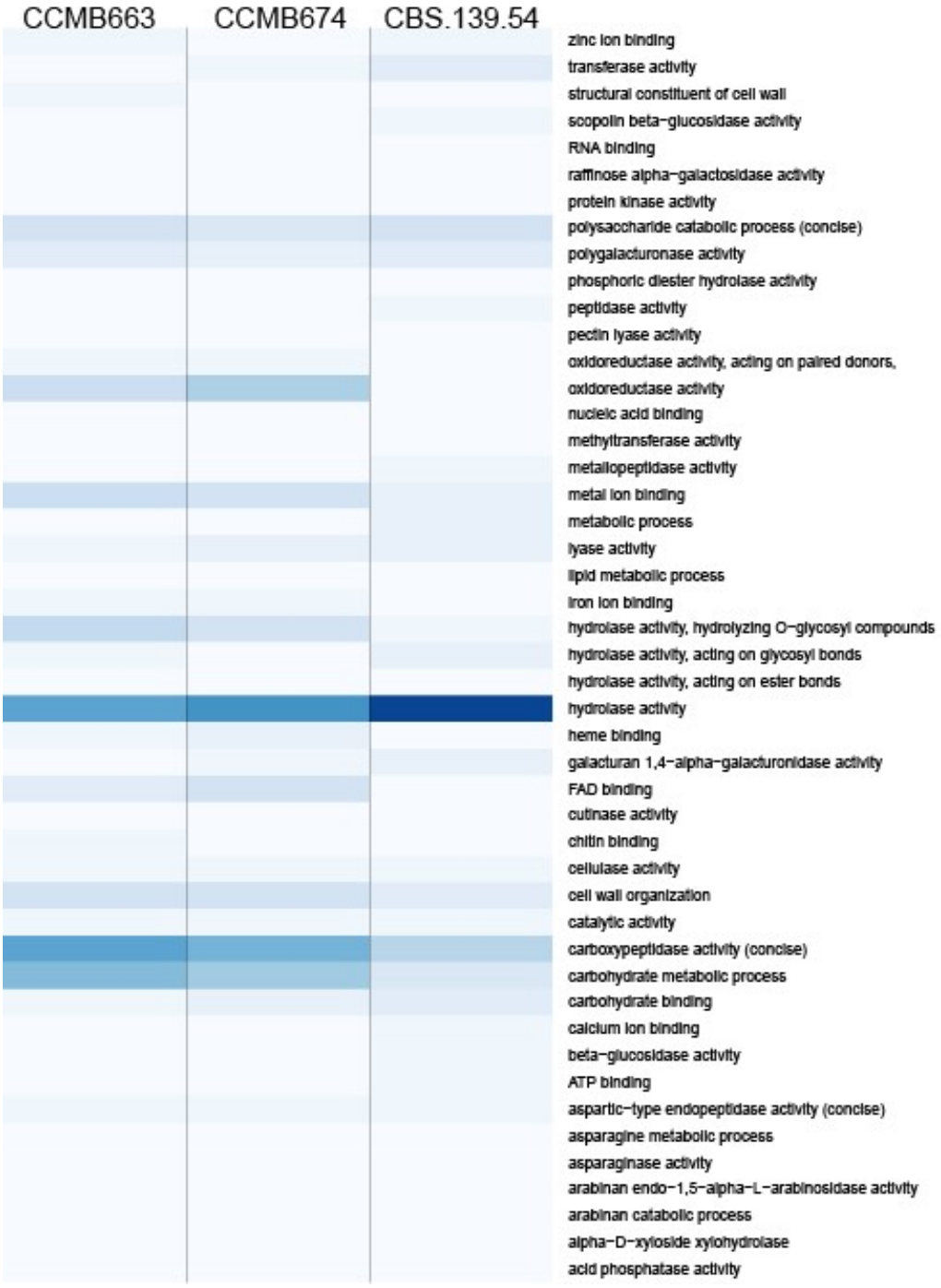


**Figure 9:** Heatmap with Gene Ontology terms (rows) shared by CCMB663 and CCMB674, or by each of these with CBS 139.54 (columns). Darker colors indicate higher abundance. Blank spaces mean that function was not detected in that specific isolate.

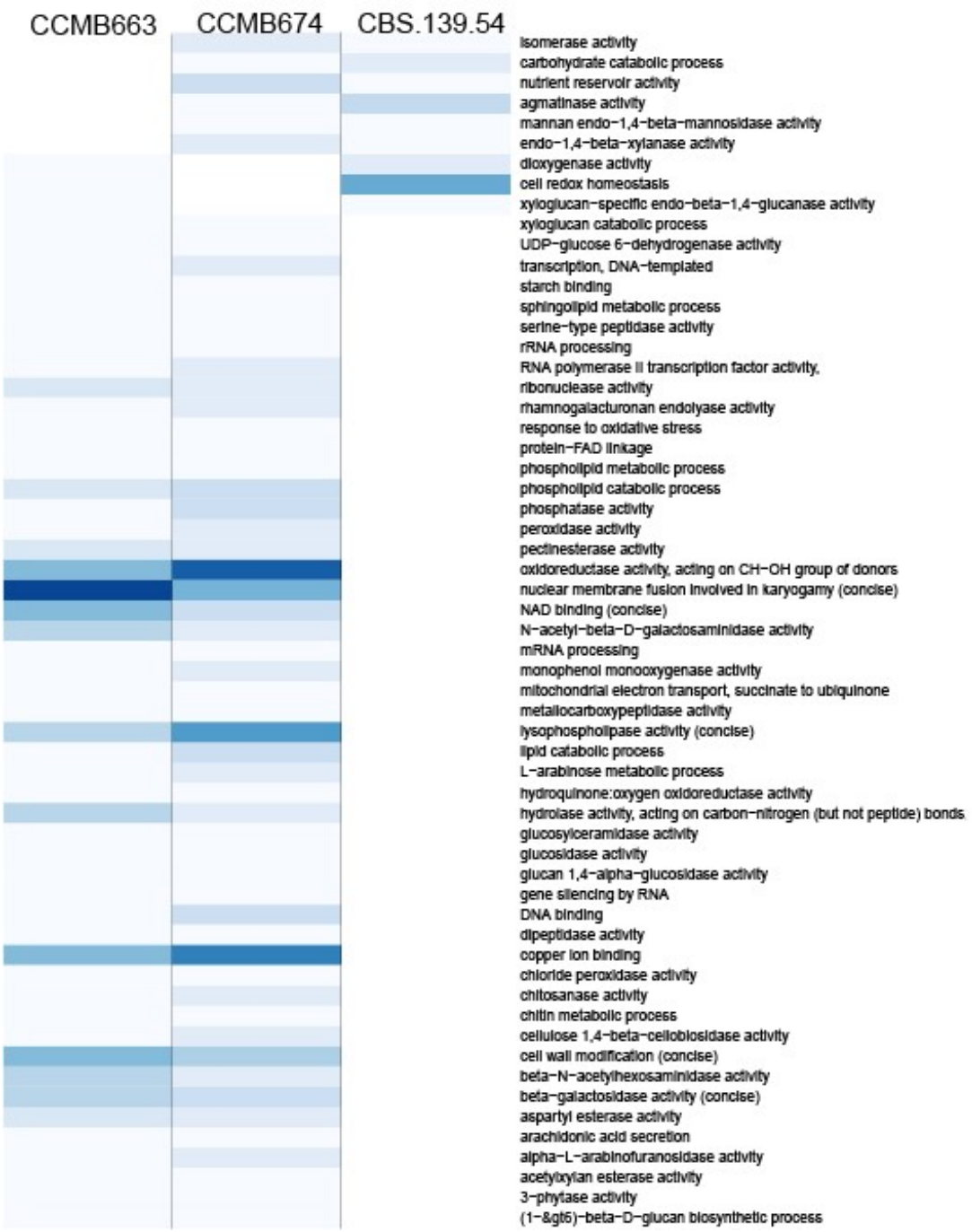


**Table 4:** Exclusive GO term distribution according to associated proteins in CCMB663, CCMB674 and CBS 139.54.

| Isolate | GO term | Term Description | Protein Name | E. C. number |
|---|---|---|---|---|
| AW 674 | GO:0008496 | mannan endo-1,6-alpha-mannosidase activity | mannan endo-1,6-alpha-mannosidase | 3.2.1.101 |
| AW 674 | GO:0009116 | nucleoside metabolic process | kinesin light chain | Unavailable |
| AW 674 | GO:0003735 | structural constituent of ribosome | 60S acidic ribosomal protein P2 | Unavailable |
| AW 674 | GO:0019863 | IgE binding | allergen Asp F4 | Unavailable |
| | | | allergen Asp F4 | Unavailable |
| AW 674 | GO:0102131 | 3-oxo-glutaryl | short chain dehydrogenase/reductase family | 1. -. -.- |
| AW 674 | GO:0008199 | ferric iron binding | extracellular dioxygenase | 1.13.11.- |
| | | | GPI anchored dioxygenase | 1.13.11.- |
| AW 674 | GO:0033897 | ribonuclease T2 activity | ribonuclease T2 | 4.6.1.19 |
| AW 674 | GO:0004497 | monooxygenase activity | cytochrome P450 | 1.14 -.- |
| | | | cytochrome P450 | 1.14.-.- |
| | | | cytochrome P450 | 1.14.-.- |
| AW 674 | GO:0005840 | ribosome | 60S acidic ribosomal protein P2 | Unavailable |
| AW 674 | GO:0006529 | asparagine biosynthetic process | asparagine synthase related protein | 6.3.5.4 |
| AW 674 | GO:0004725 | protein tyrosine phosphatase activity | 3-phytase B | 3.1.3.8 |
| AW 674 | GO:0043333 | 2-octaprenyl-6-methoxy-1,4-benzoquinone methyltransferase activity | beta-mannosidase A | 3.2.1.25 |
| AW 674 | GO:0030328 | prenylcysteine catabolic process | Prenylcysteine oxidase | 1.8.3.5 |
| AW 674 | GO:0042744 | hydrogen peroxide catabolic process | catalase R | 1.11.1.6 |

| | | | | |
|---|---|---|---|---|
| AW 674 | GO:0031314 | extrinsic component of mitochondrial inner membrane | beta-mannosidase A | 3.2.1.25 |
| AW 674 | GO:0006744 | ubiquinone biosynthetic process | beta-mannosidase A | 3.2.1.25 |
| AW 674 | GO:0006414 | translational elongation | 60S acidic ribosomal protein P2 | Unavailable |
| AW 674 | GO:0004316 | 3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity | short chain dehydrogenase/reductase family | 1. -. -.- |
| AW 674 | GO:0004096 | catalase activity | catalase R | 1.11.1.6 |
| AW 674 | GO:0000225 | N-acetylglucosaminylphosphatidylinositol deacetylase activity | F5/8 type C domain protein | Unavailable |
| AW 674 | GO:0016407 | acetyltransferase activity | Sister chromatid cohesion acetyltransferase Eco1 | 2.3.1.- |
| AW 674 | GO:0000324 | fungal-type vacuole | RNA Polymerase II CTD phosphatase Fcp1 | 3.1.3.16 |
| AW 674 | GO:0007062 | sister chromatid cohesion (=GO:0000070) | Sister chromatid cohesion acetyltransferase Eco1 | 2.3.1.- |
| AW 674 | GO:0006308 | DNA catabolic process | S1/P1 nuclease | 3.1.30.1 |
| AW 674 | GO:0034480 | phosphatidylcholine phospholipase C activity | phosphoesterase | 3.1.3.2 |
| AW 674 | GO:0016702 | oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen | extracellular dioxygenase | 1.13.11.- |
| | | | GPI anchored dioxygenase | 1.13.11.- |
| | | | GPI anchored dioxygenase | 1.13.11.- |
| AW 674 | GO:0030596 | alpha-L-rhamnosidase activity | alpha-L-rhamnosidase C | 3.2.1.40 |
| AW 674 | GO:0043324 | pigment metabolic process involved in developmental pigmentation | multicopper oxidase | 1. -. -.- |

| | | | | |
|---|---|---|---|---|
| AW 674 | GO:0045132 | meiotic chromosome segregation (=GO:0000070) | Sister chromatid cohesion acetyltransferase Eco1 | 2.3.1.- |
| AW 674 | GO:0009251 | glucan catabolic process | endoglucanase A | 3.2.1.6 |
| AW 674 | GO:0001735 | prenylcysteine oxidase activity | Prenylcysteine oxidase | 1.8.3.5 |
| AW 674 | GO:0042578 | phosphoric ester hydrolase activity | SacI domain protein | Unavailable |
| AW 674 | GO:0042803 | protein homodimerization activity | feruloyl esterase B-1 | 3.1.1.73 |
| AW 674 | GO:0005773 | vacuole | carboxypeptidase Y | 3. 4.16.- |
| AW 674 | GO:0070940 | dephosphorylation of RNA polymerase II C-terminal domain | RNA Polymerase II CTD phosphatase Fcp1 | 3.1.3.16 |
| AW 674 | GO:0004459 | L-lactate dehydrogenase activity | l-lactate dehydrogenase | 1.1.1.27 |
| AW 674 | GO:0006506 | GPI anchor biosynthetic process | F5/8 type C domain protein | Unavailable |
| AW 674 | GO:0006725 | cellular aromatic compound metabolic process | extracellular dioxygenase | 1.13.11.- |
| | | | GPI anchored dioxygenase | 1.13.11.- |
| AW 674 | GO:0102132 | 3-oxo-pimeloyl-[acp] methyl ester reductase activity | short chain dehydrogenase/reductase family | 1. -. -.- |
| AW 674 | GO:0032366 | intracellular sterol transport | ML domain protein | Unavailable |
| AW 674 | GO:0033609 | oxalate metabolic process | oxalate decarboxylase oxdC | 4.1.1.2 |
| | | | oxalate decarboxylase | 4.1.1.3 |
| AW 674 | GO:0004359 | glutaminase activity | glutaminase | 3.5.1.2 |
| AW 674 | GO:0004519 | endonuclease activity | S1/P1 nuclease | 3.1.30.1 |
| AW 674 | GO:0030600 | feruloyl esterase activity | feruloyl esterase B-1 | 3.1.1.73 |

| | | | | |
|---|---|---|---|---|
| AW 674 | GO:0008420 | CTD phosphatase activity | RNA Polymerase II CTD phosphatase Fcp1 | 3.1.3.16 |
| AW 674 | GO:0030248 | cellulose binding | 1,4-beta-D-glucan cellobiohydrolase C | 3.2.1.91 |
| | | | endo-beta-1,4-glucanase D | 3.2.1.6 |
| | | | endoglucanase A | 3.2.1.6 |
| | | | family 61 glycosyl hydrolase | 3.2.1.- |
| AW 674 | GO:0000070 | mitotic sister chromatid segregation | Sister chromatid cohesion acetyltransferase Eco1 | 2.3.1.- |
| AW 674 | GO:0016311 | dephosphorylation | 3-phytase B | 3.1.3.8 |
| AW 674 | GO:0006890 | retrograde vesicle-mediated transport, Golgi to ER | protein transport protein Sec39 | Unavailable |
| AW 674 | GO:0019752 | carboxylic acid metabolic process | l-lactate dehydrogenase | 1.1.1.27 |
| AW 674 | GO:0004322 | ferroxidase activity | multicopper oxidase | 1. -. -.- |
| AW 674 | GO:0046564 | oxalate decarboxylase activity | oxalate decarboxylase | 4.1.1.2 |
| AW 674 | GO:0004066 | asparagine synthase (glutamine-hydrolyzing) activity | asparagine synthase related protein | 6.3.5.4 |
| AW 663 | GO:0006801 | superoxide metabolic process | cytosolic Cu/Zn superoxide dismutase | 1.15.1.1 |
| AW 663 | GO:0016872 | intramolecular lyase activity | Six-hairpin glycosidase | 3. 2. 1.- |
| AW 663 | GO:0047682 | aryl-alcohol oxidase activity | glucose-methanol-choline oxidoreductase | 1.1.99.10 |
| AW 663 | GO:0043864 | indoleacetamide hydrolase activity | amidase | 3.5.1.4 |
| AW 663 | GO:0047739 | cephalosporin-C deacetylase activity | carboxylesterase | 3.1.1.1 |
| AW 663 | GO:0030570 | pectate lyase activity | pectate lyase | 4.2.2.2 |
| AW 663 | GO:0031408 | oxylipin biosynthetic process | heme peroxidase | 1.11. 1.- |

| | | | | |
|---|---|---|---|---|
| AW 663 | GO:0004040 | amidase activity | glucose-methanol-choline oxidoreductase | 1.1.99.10 |
| AW 663 | GO:0000155 | phosphorelay sensor kinase activity | histidine kinase | 2.7.13.3 |
| AW 663 | GO:0019748 | secondary metabolic process | FAD-binding domain-containing protein | Unavailable |
| AW 663 | GO:0052761 | exo-1,4-beta-D-glucosaminidase activity | glycoside hydrolase | 3.2.1.- |
| AW 663 | GO:0008239 | dipeptidyl-peptidase activity | alpha/beta-hydrolase | 3. -. -.- |
| AW 663 | GO:0000009 | alpha-1,6-mannosyltransferase activity | glycosyl transferase | 2. 4. -.- |
| | | | alpha-1,6-mannosyltransferase subunit | 2. 4. 1.- |
| AW 663 | GO:0047646 | alkanal monooxygenase (FMN-linked) activity | cytochrome P450 | 1.14.-.- |
| AW 663 | GO:0006508 | proteolysis | endoprotease endo-Pro | 3.4.21.26 |
| AW 139 | GO:0004556 | alpha-amylase activity | alpha-amylase | 3.2.1.1 |
| | | | alpha-amylase | 3.2.1.1 |
| AW 139 | GO:0004521 | endoribonuclease activity | ribonuclease-domain-containing protein | Unavailable |
| AW 139 | GO:0003756 | protein disulfide isomerase activity | disulfide isomerase TigA | 5.3.4.1 |
| | | | thioredoxin-domain-containing protein | Unavailable |
| | | | protein disulfide isomerase | 5.3.4.1 |
| | | | thioredoxin-domain-containing protein | Unavailable |
| | | | thioredoxin-domain-containing protein | Unavailable |
| AW 139 | GO:0006458 | *de novo* protein folding | ROT1 | Unavailable |
| AW 139 | GO:0000287 | magnesium ion binding | phosphoglucomutase PgmA | 5.4.2.2 |
| AW 139 | GO:0004674 | protein serine/threonine kinase activity | cAMP-dependent protein kinase | 2.7.11.11 |

| | | | | |
|---|---|---|---|---|
| AW 139 | GO:0006164 | purine nucleotide biosynthetic process | bifunctional purine biosynthesis protein Ade16 | 3.5.4.10 |
| AW 139 | GO:0103025 | alpha-amylase activity (releasing maltohexaose) | alpha-amylase | 3.2.1.1 |
| AW 139 | GO:0006080 | substituted mannan metabolic process | mannan endo-1,4-beta-mannosidase A | 3.2.1.78 |
| AW 139 | GO:0003861 | 3-isopropylmalate dehydratase activity | 3-isopropylmalate dehydratase | 4.2.1.33 |
| AW 139 | GO:0016868 | intramolecular transferase activity, phosphotransferases | phosphoglucomutase PgmA | 5.4.2.2 |

### 5.4 Effector Analysis

By further analyzing our predicted secretome for isolates CCMB663, CCMB674, and CBS 139.54, we identified various effectors for each isolate. The vast majority of those small proteins did not present similarity with any known protein on PHIB-blast.

**Table 5:** Details on the selected candidate effectors on CCMB674, CCMB 663 and CBS 139.54. Columns indicate the isolate to which proteins belong, followed by their names, detected Pfam domain (if any) and its region in the peptide. Last column represents search ids of PHI-blast hits.

| Isolate | Protein | Pfam | Region | PHI-base hits |
|---|---|---|---|---|
| CCMB663 | protein_987 | Antifungal protein | 37-86 | 5374 |
| CCMB663 | protein_1038 | *none* | *none* | 2452 |
| CCMB663 | protein_2879 | Pectate lyase | 68-270 | 7940 |
| CCMB663 | protein_4449 | FKBP | 33-125 | 2305, 548, 4584, 4583 |
| CCMB663 | protein_10801 | Pectate lyase | 68-270 | 3226 |
| CCMB663 | protein_11811 | Peroxidase 2 | 14-242 | 5188 |
| CCMB663 | protein_11861 | *none* | *none* | 7703, 4909 |
| CCMB674 | protein_908 | FKBP | 33-125 | 548, 2305, 4584, 4583 |
| CCMB674 | protein_3004 | Hydrophobin | 43-142 | 458, 291 |
| CCMB674 | protein_3511 | Antifungal protein | 37-86 | 5374 |
| CCMB674 | protein_6718 | Peroxidase 2 | 14-242 | 5188 |
| CCMB674 | protein_8317 | Hydrophobin | 43-142 | 7940 |
| CCMB674 | protein_9381 | Pectate lyase | 68-270 | 3226, 4618, 4619, 4617, 6855, 6854, 6856, 222 |
| CCMB674 | protein_10839 | LysM | 52-91<br>128-167<br>198-237 | 7376, 6832 7381, 7665, 6831, 737, 6833, 7380 7379 |
| CBS 139.54 | protein_782 | FKBP | 33-125 | 548, 2305, 4584, 4583 |
| CBS 139.54 | protein_4387 | Thioredoxin | 20-127 | 2644<br>7258 |
| CBS 139.54 | protein_9954 | Hydrophobin | 59-158 | 7940 |
| CBS 139.54 | protein_11329 | Alpha-amylase | 65-178 | 3904, 6732, 2504, 3903 |
| CBS 139.54 | protein_13639 | Glucanosyltransferase | 16-303 | 33, 6713, 522, 2808, 6714, 6712, 6715, 434, 6711, 6530 |

In total, we identified 66 candidate effectors, seven of which were similar to known effectors on CCMB663, 75 proteins in CCMB674, out of which seven presented hits in the database, and 36 in CBS 139.54, out of which 5 could be further identified with PHIB-blast. All of the further characterized effectors exhibited predicted disordered regions, except for the antifungal proteins in CCMB663 and CCMB674.

Only one protein with no recognizable protein domains, protein_11861, displayed a predicted binding site region, in a highly disordered region. Proteins with the same Pfam domain were highly similar (identity >98% with BLASTP), except for protein_9381 in CCMB663, which did not share any similarity with the other pectate lyases, and protein_3004 in CCMB674, that shares only a 12 amino acid region with the other two hydrophobins.

**Figure 10:** Disordered regions and binding site predictions for *Aspergillus welwitschiae* CCMB663 proteins. **A:** protein_987 **B:** protein_1038 **C:** protein_2879 **D:** protein_4449 **E:** protein_10801 **F:** protein_11811 **G:** protein_11861

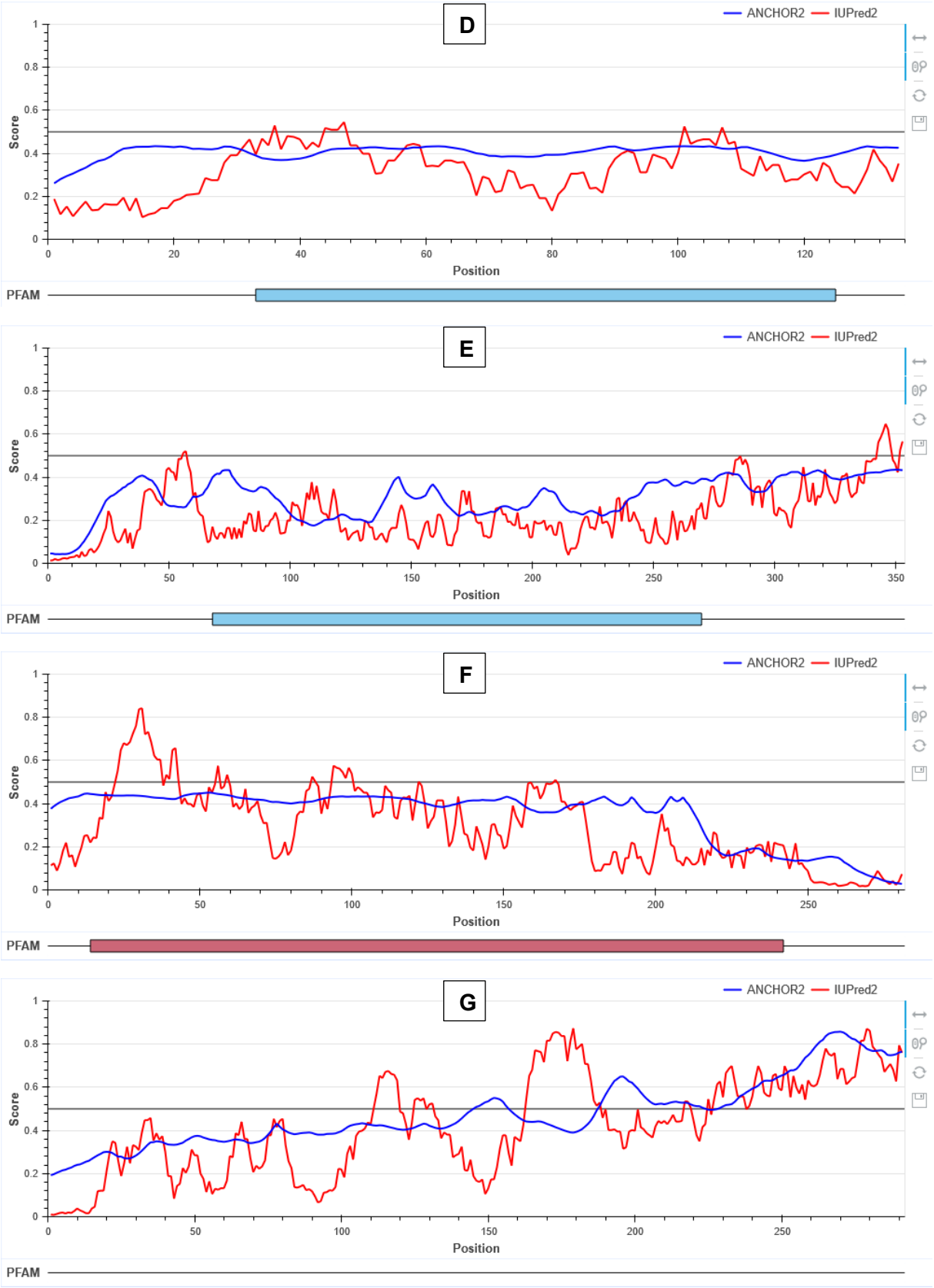
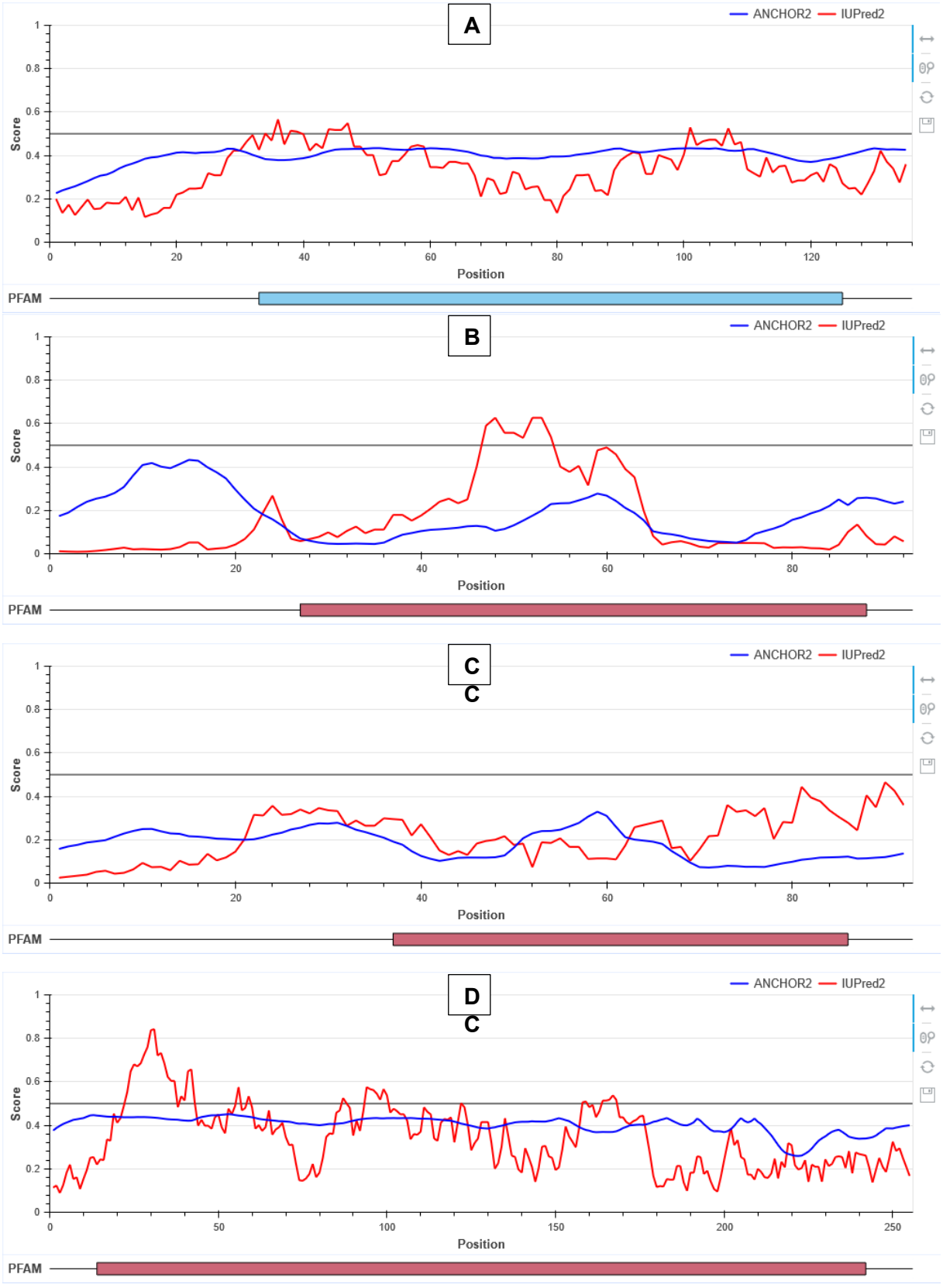

D
ANCHOR2
IUPred2
Score
Position
PFAM
E
ANCHOR2
IUPred2
Score
Position
PFAM
F
ANCHOR2
IUPred2
Score
Position
PFAM
G
ANCHOR2
IUPred2
Score
Position
PFAM

**Figure 11:** Disordered regions and binding site predictions for *Aspergillus welwitschiae* CCMB674 proteins. **A:** protein_908 **B:** protein_3004 **C:** protein_3511 **D:** protein_6718 **E:** protein_8317 **F:** protein_9381 **G:** protein_10839



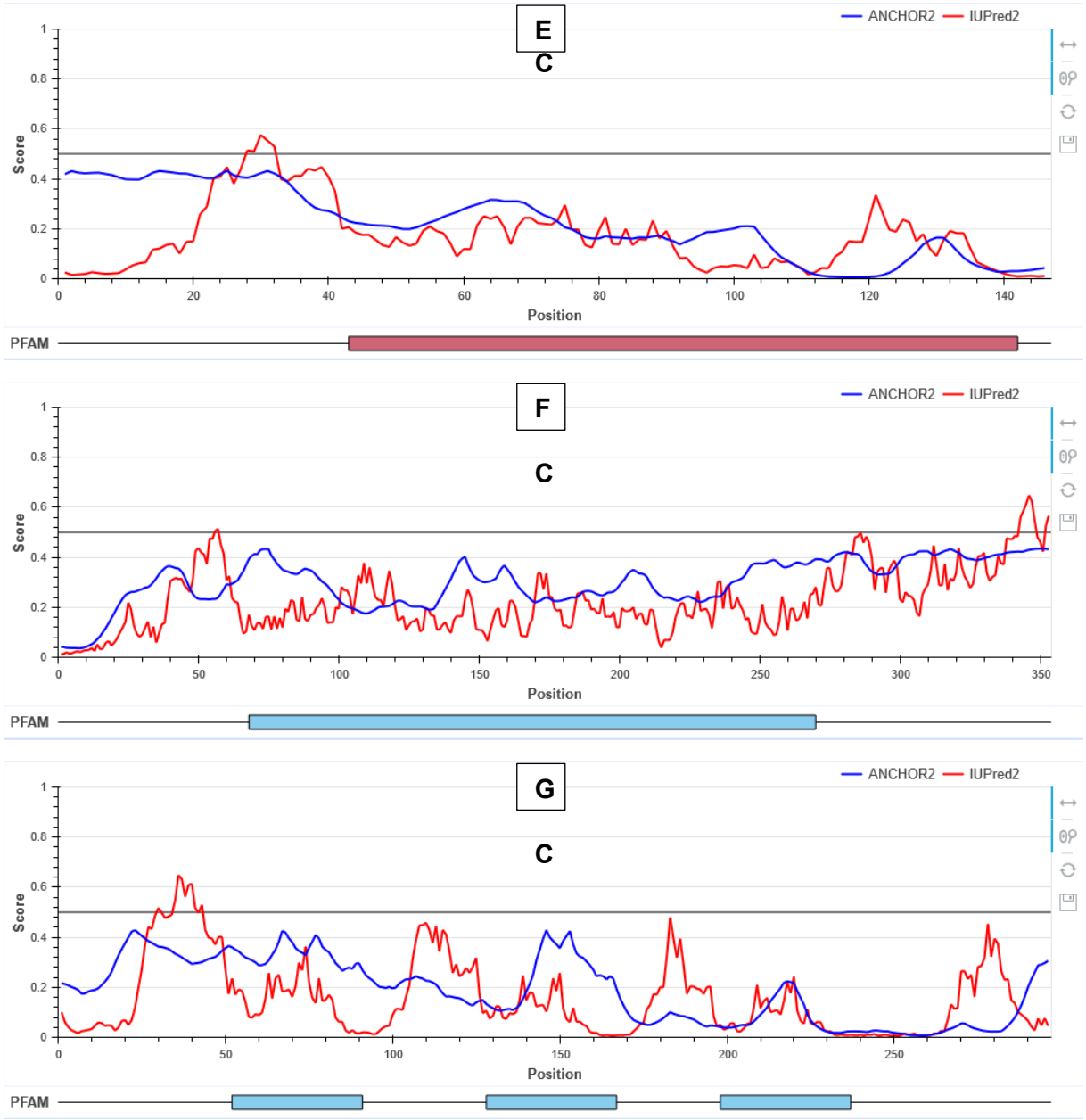
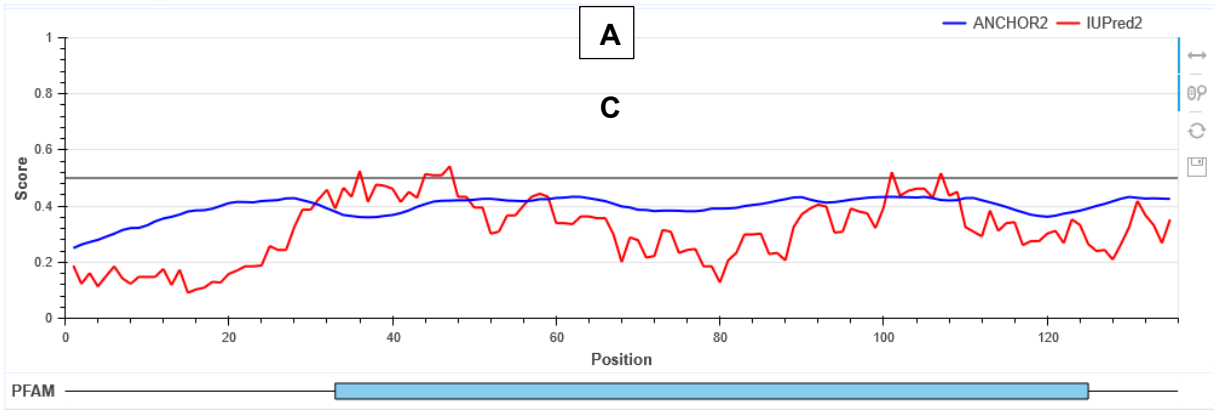
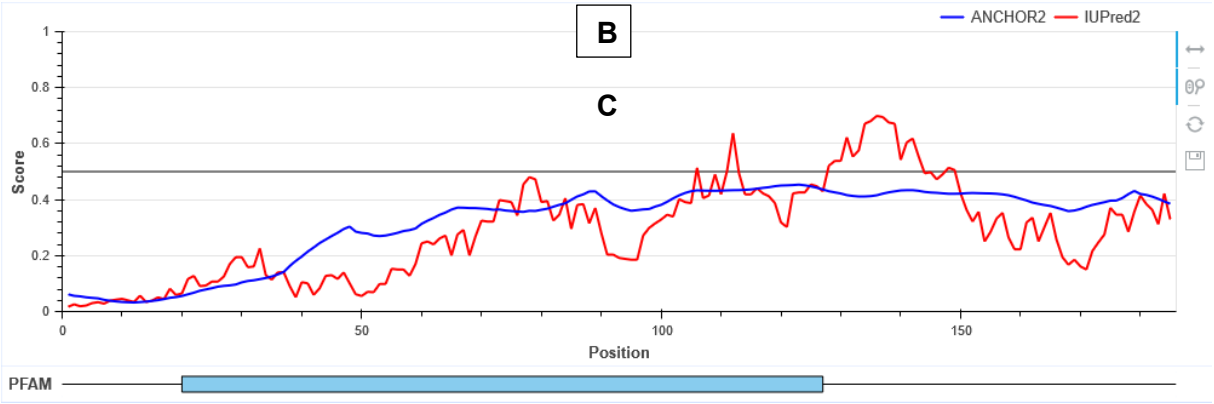
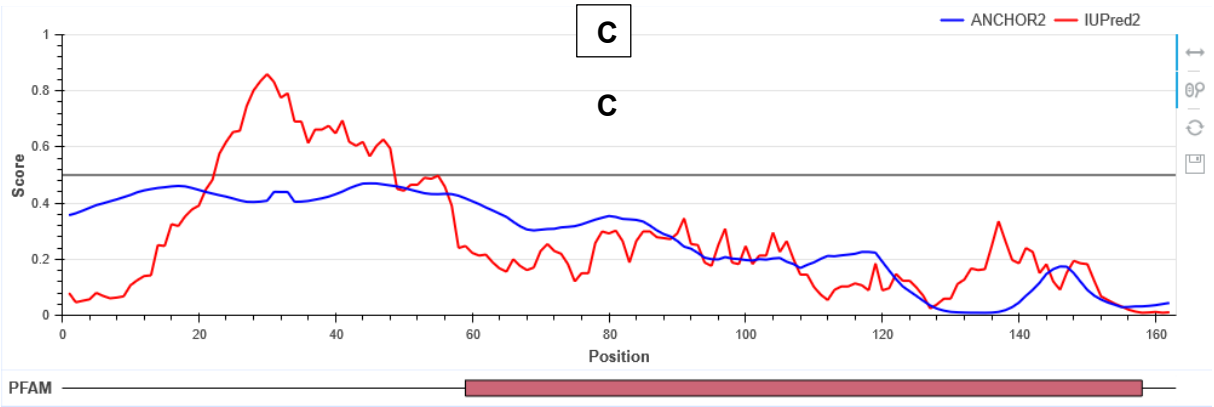
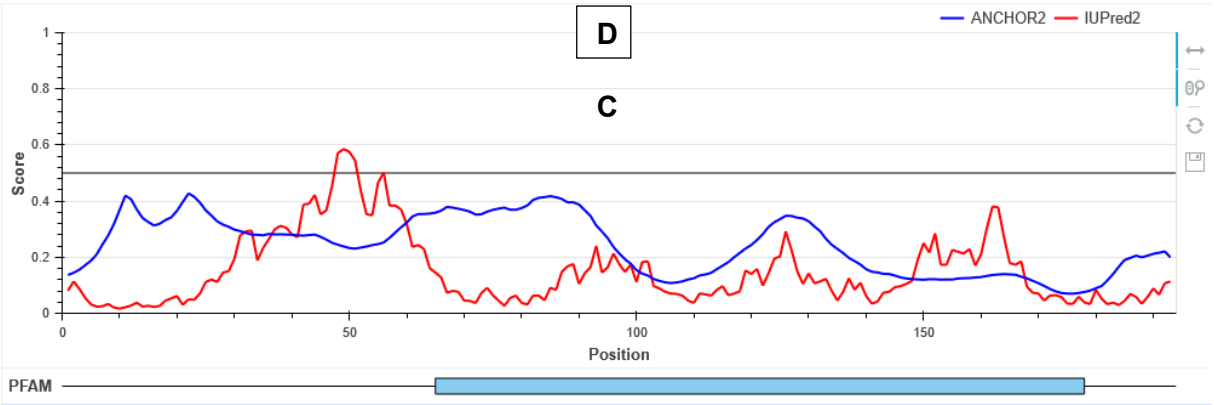


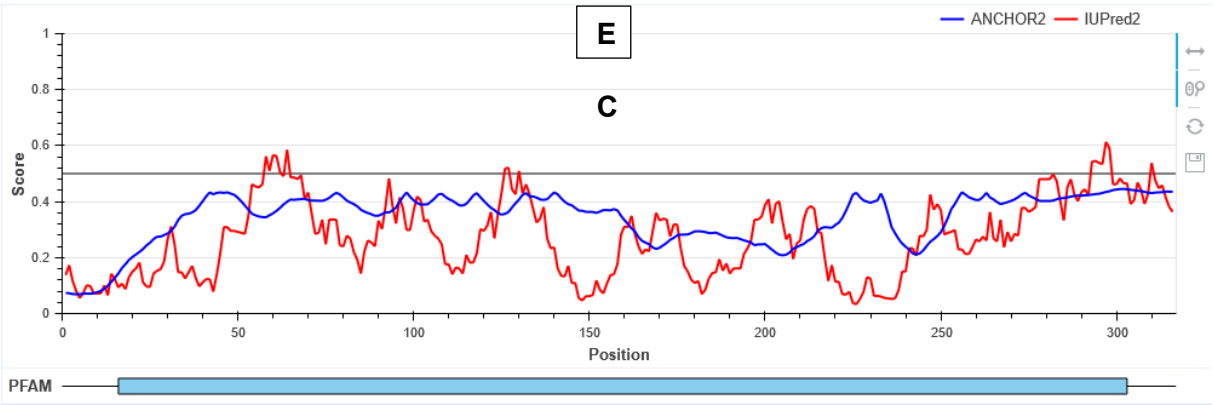

**Figure 12:** Disordered regions and binding site predictions for *Aspergillus welwitschiae* CBS 139.54 proteins. **A:** protein_782 **B:** protein_4387 **C:** protein_9954 **D:** protein_11329 **E:** protein_13639




B
C
ANCHOR2
IUPred2
Score
Position
PFAM


C
C
ANCHOR2
IUPred2
Score
Position
PFAM


D
C
ANCHOR2
IUPred2
Score
Position
PFAM


E
C
ANCHOR2
IUPred2
Score
Position
PFAM

### 5.5 CAZyme Analysis

Our CAZyme annotation approach was adapted from Zhao (2013). There are different families of carbohydrate degrading enzymes, related to different steps of lignin and carbohydrate degradation. Figure 13 shows the distribution of most abundant CAZymes in our three isolates.

**Table 6:** Percentual relative distribution of CAZymes (columns) in three isolates of *Aspergillus welwitschiae* (rows).

| | Glycoside Hydrolases | Glycosyl Transferases | Auxiliary Activities | Carbohydrate Esterases | Carbohydate-Binding Modules | Poly-saccharide Lyases |
|---|---|---|---|---|---|---|
| CCMB663 | 47.39 | 18.98 | 13.57 | 8.50 | 8.64 | 2.88 |
| CCMB674 | 45.12 | 19.80 | 16.52 | 9.75 | 6.37 | 2.41 |
| CBS 139.54 | 48.00 | 24.12 | 12.78 | 7.86 | 5.85 | 1.37 |

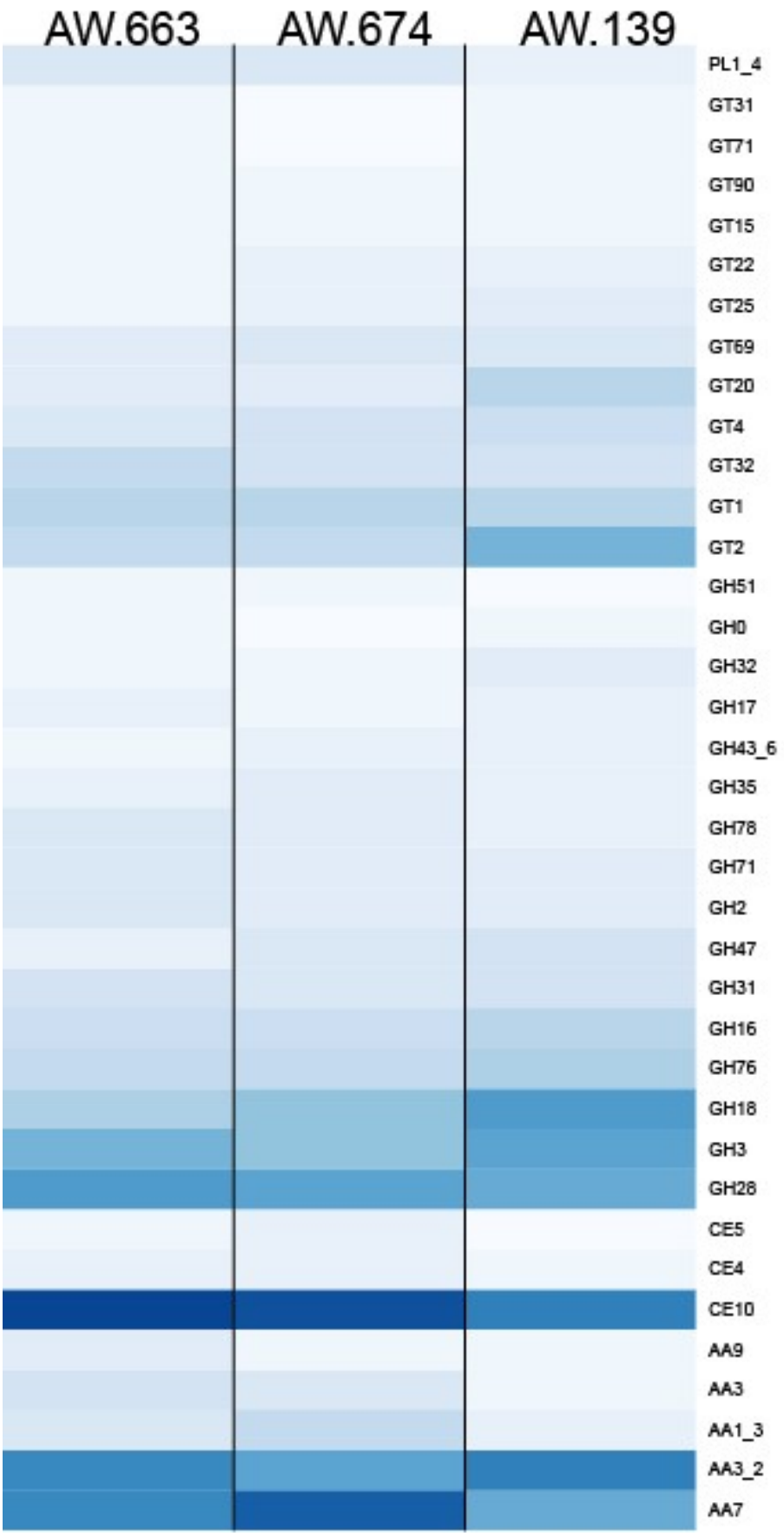
Our Carbohydrate-Active Enzyme annotation workflow resulted in a very similar CAZyme content for all three isolates, with almost half of the proteins represented by glycoside hydrolases (GH), followed by approximately 20% of glycosyltransferases (GT), 15% of auxiliary activities (AA) and less than 15% for all carbohydrate esterases (CE), carbohydrate-binding modules (CBM) and polyssacharide lyases (PL). The glycoside hydrolases (GH) class was divided into 153 families based in the CAZy database. Among those, 68 were detected in the two sisal isolates and CBS 139.54. A total of 56 families were present in all three *A. welwitschiae* isolates, six families in any two genomes and five families in only one genome. GlycosylTransferases (GT) are classified into 106 families, among which 36 were detected in our three isolates. A total of 34 of those enzymes occurred in all the three *A. welwitschiae* isolates and only two families detected in two isolates. Auxiliary Activities (AA) are classified into 15 families. Nine of those families were detected in our three isolates, and no exclusive-occurring protein family was detected. Carbohydrate Esterases (CE) are classified into 16 families, 12 of which were detected in total. A total of 11 families were present in all three isolates and only one family in one isolate. Among 88 families of Carbohydrate-Binding Modules (CBMs), 14 were detected in all isolates. A total of 11 of those enzymes were present in all three genomes, two families were detected in two isolates, and one family in only one isolate. Polysaccharide Lyases (PLs) are classified into 28 families, of which only five were detected in the isolates, shared by all three of them. In total, out of the 406 CAZyme families from the CAZy database, 144 were identified

in our three isolates of *Aspergillus welwitschiae*. Except for minor differences, the relative frequency of the most abundant (equal or more than 1%) predicted CAZymes families in each isolate were quite similar. Table 7 depicts the distribution and relative frequency in the three *A. welwitschiae* isolates.

**Table 7:** Distribution of CAZymes (rows) into all three isolates or specific conditions (columns).

| CAZyme type | All Three Isolates | One or Two Isolates |
|---|---|---|
| Glycoside hydrolases | GH0 GH1 GH10 GH105 GH106 GH11 GH12 GH127 GH128 GH13 GH132 GH133 GH135 GH145 GH15 GH16 GH17 GH18 GH19 GH2 GH20 GH27 GH28 GH29 GH3 GH30 GH31 GH32 GH33 GH35 GH36 GH37 GH38 GH43 GH47 GH5 GH51 GH53 GH54 GH55 GH6 GH62 GH63 GH65 GH67 GH7 GH71 GH72 GH75 GH76 GH78 GH79 GH88 GH92 GH93 | (CCMB663 and CBS 139.54) GH74 GH81 GH109 GH125 GH131<br>(CCMB674 and CBS 139.54) GH139<br>(CBS 139.54) GH4 GH26 GH134 GH142 GH140 |
| Glycosyl Transferases | GT0 GT0+GT7 GT1 GT15 GT2 GT20 GT20+GT20 GT22 GT24 GT25 GT3 GT30 GT31 GT32 GT33 GT34 GT39 GT4 GT47 GT48 GT50 GT55 GT57 GT58 GT59 GT62 GT66 GT69 GT71 GT76 GT8 GT90 GT90+GT8 | (CCMB663 and CBS 139.54) GT35 GT41 |
| Auxiliary Activities | AA1 AA2 AA3 AA4 AA6 AA7 AA8 AA9 AA11 | |
| Poly-saccharide Lyases | PL1_4 PL1_7 PL27 PL4_1 PL4_3 | |
| Carbo-hydrate-Binding Modules | CBM1 CBM13 CBM18 CBM20 CBM24 CBM42 CBM43 CBM48 CBM50 CBM63 CBM67 | (CCMB663 and CBS 139.54) CBM21, CBM35<br>(CBS 139.54) CBM38 |
| Carbo-hydrate Esterases | CE0 CE1 CE10 CE12 CE16 CE2 CE3 CE4 CE5 CE8 CE9 | (CBS 139.54) CE7 |

**Figure 13:** Relative abundance of CAZymes (rows) in CCMB663, CCMB674 and CBS 139.54 (columns). Darker colors indicate higher abundance.



## 6 DISCUSSION

In this study, we provided two new genomes for *A. welwitschiae* isolates and analyzed its content, looking for insights on how these fungi infect and colonize its host, *Agave sisalana*. Until the completion of this work, the only complete genome available for *A. welwitschiae* was that of isolate CBS 139.54 (VESTH et al., 2018), which was originally isolated from *Welwitschia mirabilis*, thus naming the species. We used this genome for comparison purposes in all our analyses. Our assembly quality control results showed these three isolates (*Aspergillus welwitschiae* CCMB663*, Aspergillus welwitschiae* CCMB674, *and Aspergillus welwitschiae* CBS 139.54) were different among themselves in secreted protein content, however, those differences seemed to follow the haplotype characterization for those species. CCMB674, which belongs to Haplotype 6 that is endemic to the sisal-producing region of the State of Bahia (DUARTE et al., 2018), was farther different from the other two isolates, both of which belong to Haplotype 1. Results also suggest CCMB674 was different in gene content, as seen in Figure 4. *Aspergillus welwitschiae* is a broad-host necrotrophic fungus, that acts as an opportunistic pathogen. Moore et al. (2011) mention different classic characteristics of necrotrophic fungi that summarizes many attributes found in *Aspergillus welwitschiae*. As mentioned in our introduction, this fungus usually attacks damaged plants, entering via wounds. Once inside the host tissues, *A. welwitschiae* grows intercellularly and intracellularly through dead cells, which are rapidly killed through copious secreted cell-wall-degrading (CWDE) enzymes and effectors, which can be observed in our results, with several hydrolase-related gene functions and pectate lyases identified in our effector proteins, for an example. These enzymes act on the disruption of the plant cell wall, and the sub-products of this degradation act as elicitors of the plant defenses.

Even though we expect few differences in genome sequences for same-species isolates in comparison with broader classifications, such as genus or family levels, our differences match both the plasticity in genome size observed in the Nigri section of the genus Aspergillus (VESTH et al., 2018), and also the plasticity in genome size for different isolates of the closest species, *Aspergillus niger*, according to the genomes available in GenBank (https://www.ncbi.nlm.nih.gov/genome/genomes/429), which varies from 33.825 Mbp up to 36.4508 Mbp. It is also worth noting that, according to results described by Patel et al (2018), different assembly strategies might result in

different results, which might also explain some of those differences among the analyzed isolates.

As for the mitochondrial genomes, all three *A. welwitschiae* isolates possess essentially a similar gene content, corroborating with results obtained by Juhász (2008), with differences limited to a translocation in NAD5 and a split in RPS5 in CBS 139.54, and a duplication in NAD4l in this genome and CCMB663, which is not present in CCMB674 nor in *Aspergillus niger* N909. This genome is also different in homing endonuclease (HE) content, considering it contains a GIY-YIG region, which is not present in *A. welwitschiae*, which varies only in LAGLIDADG content. HE are small proteins that recognize and confer mobility to specific regions of a genome (BELFORT; BONOCORA, 2008) analogously to restriction enzymes.

As we concluded the assembly and annotation of our two fungal isolates, CCMB663 and CCMB674, we worked on our main goal, which is to describe its relationship with the host (*Agave sisalana*, or *Welwitschia mirabilis* for CBS 139.54). In the literature, *A. welwitschiae* is described as a necrotrophy or saprotroph (feeding on dead organic matter), acting sporadically as an opportunistic pathogen. As mentioned previously, in the sisal bole rot, *A. welwitschiae* invades the plant through wounds in the leaf base, as a consequence of leaf excision for fiber extraction. This means that the fungus initially feeds on the dead leaf base tissue, facing the living bole tissue (and the plant's stress response) at some point during colonization (DUARTE et al., 2018). Therefore, we considered secretomics as a key perspective to study the fungus-plant relationship. The secretome is the totality of proteins secreted by a cell or an organism to the extracellular environment (FÉLIX et al., 2016) including or not proteins attached to the cell surface (GIRARD et al., 2013). Secreted proteins may include many different proteins associated with pathogenicity since those proteins will be directly in contact with the host environment (WARTENBERG et al., 2011), such as effectors, and carbon, lipid, and protein degradation enzymes, considering that fungi process their nutrition externally, before internalizing smaller, less complex molecules. Our secretome results seemed to be related to the geographic distribution and haplotyping of our three isolates; Those three isolates are isolated geographically, one of them belonging to a whole different haplotype (CCMB674, in Haplotype 6). Thus, even though those isolates belong to the same species, reflecting on 55% of the secretome being shared by all three isolates or any two out of these three, 45% of its secretome is unique to CCMB663, CCMB674 or CBS 139.54. It is also worth noting that

CCMB674 has a smaller gene content according to out BUSCO results, and a smaller protein set predicted on MAKER. Nevertheless, this does not reflect on its secretome content, which is greater than those of the other two isolates, including 612 proteins against 461 proteins in CCMB663 and 429 in CBS 139.54.

We found that shared Gene Ontology terms tend to follow a profile of predominantly lipid, protein, and carbon degradation-related terms, the latter more abundantly and commonly describing functions associated with cell wall-degrading enzymes (CWDE), and genes related to the fungus internal metabolism, in a smaller scale. This trend was observed in all the three isolates, as seen in Figure 8, and also in two out of the three fungal isolates (those mostly shared by CCMB663 and CCMB674 – Figure 9). The carbon degrading biological processes and molecular functions describe proteins acting on different carbohydrates, such as galacturan, glucose, galactose, raffinose, chitin, cellulose, and arabinan. Most of those carbohydrates are constituents of the cell wall (except chitin, typical of the fungal cell wall) (KAVANAGH, 2011), or sub-products of its degradation, which corroborates our previous affirmation that the fungus feeds mainly on dead tissue. This includes most recurring GO terms in the three isolates, such as “hydrolase activity”. The second most recurring term, “carboxypeptidase activity” is related to protein degradation. In Figure 8 we also see some common functions and processes in the fungal metabolism, such as Zinc ion binding (Zn is a cofactor to several cellular enzymes) RNA, FAD, ATP, and heme-binding. Figure 8 also gives us an early insight on fungal pathogenicity, considering pectin lyases are common fungal virulence factors (FU et al., 2015). Our second heatmap in Figure 9 includes GO terms shared by only two of the three isolates, most of those absent in CBS 139.54. Although these terms are different from Figure 8, the profile in this heatmap is very similar, including mainly processes and functions related to carbohydrate degradation, and some functions on protein and lipid metabolism, including some of the most recurrent terms, such as “oxidoreductase activity, acting on CH−OH group of donors” and “lysophospholipase activity”, which complement terms occurring in Figure 8. In this figure, some of the most recurring terms are descriptive of the fungal metabolism, such as copper ion binding (Cu is another common enzyme cofactor).

Table 4 describes the exclusively occurring terms on CCMB663, CCMB 674 and CBS 139.54. To better describe those terms, we mapped them back to their

associated proteins and re-annotated them with BLASTP. Although these terms are unique, their description and associated proteins seem to largely follow the pattern seen in shared terms, that is, enzymes related to dead plant tissue decomposition and common metabolic processes and functions. Aside from those terms and proteins, unique functions are revealed in each isolate. In CCMB674, which contains the greater number of exclusive GO terms, one of the unique terms is “IgE binding”, returning two proteins with high similarity (95%) with *A. niger* allergen Asp F4. Even though this function is not related to virulence in any plant host, considering that plants do not possess an equivalent to antibodies, proteins causing hypersensibility reactions in humans have been reported in the three main representative species in the genus Aspergillus; *Aspergillus flavus* (HEDAYATI et al., 2007), *Aspergillus fumigatus* (CHAUDHARY; MARR, 2011) and *Aspergillus niger* (FALAHATI et al., 2016), which as previously mentioned, is the closest species to *A. welwitschiae*. Nevertheless, allergen proteins or hypersensibility reactions have never been described as caused by *A. welwitschiae*. The unique term “monooxygenase” returns three different cytochrome P450 proteins, also present in CCMB663 under the term “alkanal monooxygenase (FMN-linked)”. This is a very versatile group of proteins, which has been described as part of aflatoxin (BENNETT; KLICH; MYCOTOXINS, 2003), fumonisin (NITSCHE et al., 2012), ochratoxin metabolisms (EL KHOURY; ATOUI, 2010) and even acting on plant antimicrobial compound pisatin (STRUCK, 2006). *Aspergillus welwitschiae* is not capable of producing aflatoxin but has been described as a common producer of fumonisin and, less commonly, ochratoxin (FERRANTI et al., 2017; MASSI et al., 2016; SUSCA et al., 2016). Interestingly, very common stress defense strategies appear as exclusive GO terms on CCMB674 and CCMB663. While the first contains an extracellular catalase (cat), the latter contains an extracellular superoxide dismutase (sod) enzyme. Both of these enzymes protect the fungus aginst the plant-generated oxidative burst, which is part of the defense mechanisms. CCMB663 also contains the exclusive term “pectate lyase activity”, which is associated with and broadly studied for its effector function (EXPERT et al., 2018; FU et al., 2015). Curiously, protein_8305 was described as a non-effector protein with 92.6% confidence in our EffectorP results. This specific result deserves to be better studied.

As a part of our secretome analysis, we verified the effector proteins content in all our isolates. Effectors are small proteins that somehow facilitate the colonization of

the host by a pathogen. However, effector classification is very complex and understudied, considering the vast mechanisms of action of those proteins on the host (LORANG, 2018). A second reason is a strong characteristic of effector proteins, the presence of disordered regions. Those regions are not arranged into any type of specific secondary structure, such as alpha helices or beta sheets. Since changes in those regions do not cause effects on the protein structure, they are known to be more susceptible to mutations at the genome level, which makes its classification even more difficult. The vast majority of our candidate effectors are hypothetical proteins, as it usually happens in this type of study. Nonetheless, since we could not verify their expression through transcriptomics, as sometimes seen in the literature, we decided to analyze only those proteins with similarity with curated results on PHI-base. As a result, we found that most of the proteins with relevant hits had an identifiable domain, we could identify proteins with the same domain in different isolates that are highly identical, with only a few exceptions. The FKBP domain is present in all the three isolates, while antifungal proteins, pectate lyases and peroxidase 2 are present in CCMB663 and CCMB674, and hydrophobins were identified in CCMB674 and CBS 139.54. Other protein domains are exclusively occurring, such as a lysin motif in CCMB674 and a thioredoxin domain in CBS 139.54.

FKBP proteins appear in all the three isolates (protein_4449 in CCMB663, protein_908 in CCMB674 and protein_782 in CBS 139.54) and are highly similar among themselves. Those proteins share a similarity with the BcFKBP12 gene from the phytopathogenic fungi *Botrytis cinerea*, as described by Meléndez et al. (2009). This is a necrotroph such as *Aspergillus welwitschiae,* and this gene mediates the sensitivity of the plant host to macrolide rapamycin, through the targeting and subsequent silencing of the rapamycin signaling pathway, which mediates the response of the plant to extracellular nutrient by regulating ribosome biogenesis, translation, autophagy and the subcellular localization of transcription factors. Mélendez et al. (2009) showed that the absence of FKBP genes in *B. cinerea* decreased virulence in plant hosts such as tomato, beans, apple, and grape, but they are not essential for colonization and virulence.

Antifungal proteins, present in CCMB663 and CCMB674, represent an interesting strategy of effector proteins, which act not on the host, but rather in other, *non-self* fungi, conferring, thus a competitive advantage. According to Garrigues et al.

(2016), the identification of this type of protein is relatively rare and more common in the genera Penicillium, described by the author, and Aspergillus, to which our isolates belong. The mechanisms of action of those proteins are still largely unknown but results shown its relationship to calcium ion regulation in the target fungi, as well as compromising the cell wall stability. However, beyond its antifungal properties, Garrigues et al. (2016) also described that the deletion of antifungal proteins causes a reduction of axenic growth, abnormal hyphal morphology and delayed conidiogenesis in *Penicillium digitatum*, suggesting these proteins have more functions in cell regulation.

Pectate lyases *protein_10801* in CCMB663 and *protein_9381* in CCMB674 are similar but neither of them shares any similarity with *protein_2879* from CCMB663. Pectate Lyases also represent an intersection point between to common strategies of virulence in phytopathogenic fungi, which are effectors and cell wall degrading enzymes. Pectate lyase fits both categories because they digest cell wall components, allowing the fungi to feed on those polysaccharides but, by doing so, they also facilitate the invasion of the fungi in the plant tissues by over-stimulating the defense system to cause cell death, which is the concept of an effector protein. More specifically, pectate lyases cleave the a-1,4 glycosidic bonds of polygalacturonic acid via a b-elimination reaction. The resulting oligogalacturonates can act as defense elicitors to activate signaling pathways, activating the plant immune responses. Fu et al. (2015) described that pectate lyases play an important role in inducing cell death in pepper plants, secreted by *Phytophthora capsici.* As mentioned previously, in many effector proteins of necrotrophic fungi, the strategy relies not on silencing the plant host immune response but rather overacting it to kill the cell, so that the fungus can feed on the dead tissue.

The overexpression of the plant immune defenses produces many reactive oxygen species, in efforts to kill the fungal pathogen. This leads us to the peroxidase 2 proteins shared by CCMB663 and CCMB674 (that are highly similar). To resist the plant defenses, fungi produce many detoxifying enzymes, such as peroxidases, and also catalases and superoxide dismutases, not included here as effectors. Each of these proteins acts on a different reactive oxygen species. However, peroxidases are also considered effectors.

Hydrophobin proteins *protein_9954* in CBS 139.54 is 89.51% similar to *protein_8317* from CCMB674 (*protein_9954* contains a 16 aminoacid region absent in *protein_8317*) but both share only a low-similarity region of 12 amino acids with *protein_3004* from CCMB674. Hydrophobins are effector proteins used, among other functions, in hyphae adhesion to different surfaces. This is particularly interesting considering that necrotrophic fungi do not possess appressoria, which are specialized structures or fixation and invasion of the plant present in many biotrophic fungi (KAVANAGH, 2011). However, pathogenesis involving hydrophobins, even though not fully described, might be associated with a hydrophobic coating of fungal structures, that protect those structures from water-soluble host defenses (PARIS et al., 2003). Hydrophobins are ubiquitous in filamentous fungi, and have been described as important for fungi colonization in plant tissues for both mutualistic symbiotic fungi (GUZMÁN-GUZMÁN et al., 2017) phytopathogens (KIM et al., 2005), and also entomopathogenic fungi that live attached to plant roots in some stage of development (MOONJELY; KEYHANI; BIDOCHKA, 2018). Both of the functions mentioned previously for hydrophobins are useful in the case of *Aspergillus welwitschiae,* considering its spores are ubiquitous in the air and soil, which might contribute to its persistence in the *Agave sisalana* fields.

Other proteins did not have known protein domains, or are exclusively occurring in some of the three *Aspergillus welwitschiae* isolates, such as the lysin motif (LysM) in CCMB674 and the thioredoxin, alpha-amylase and glucanosyltransferase domains occurring in CBS 139.54 proteins.

LysM motifs are part of the “classical” effector strategies, consisting of somehow silencing the host immune response to facilitate the pathogen entrance and colonization. In plants, Lysin motifs are known to trigger the chitin-induced signaling pathway, which is a common polysaccharide in fungal cell walls (WANG et al., 2014). In fungi, proteins containing LysM avoid the triggering of these signaling pathways by binding to self chitin, outperforming the host LysM proteins (RODRIGUEZ-MORENO et al., 2018). Instead of peroxidases, present in CCMB663 and CCMB674, CBS 139.54 contains instead a thioredoxin protein, which is also involved in redox state maintenance. Thioredoxins maintain proteins disulfide reducing capacity, however, this system operates exclusively intracellularly. Extracellular thioredoxins have been described in humans (LÉVEILLARD; AÏT-ALI, 2017), a result that makes sense

considering a eukaryote organism with cells organized in tissues. According to Léveillard, thioredoxins could be maintained in the very oxidized extracellular environment by the presence of the complement system in the bloodstream. However, considering that fungi do not have specialized tissues, the presence of an extracellular thioredoxin protein remains to be described.

Finally, even though CBS 139.54 did not present pectate lyases like the other two isolates, it contains two effector proteins that are also cell wall degrading enzymes (CWDE); Alpha-amylase and Glucanosyltrasferase. This category of enzymes acts on the disruption of the cell wall and the posterior metabolism of the polysaccharides. Besides CWDE, Lorang (2018) mentions other types of effectors, two of which are present in our genomes. Necrosis-inducing-like peptides include FKBP, peroxidases 2, LysM, thioredoxins, and hydrophobins, while CWDE are represented by pectate lyases, alpha-amylase, and the glucanosyltransferase. The antifungal proteins are an exception since they do not act on the infected plant but as a possible fungal competitor.

Our results show a great diversity of Carbohydrate-active enzymes (CAZymes) in all three isolates of *Aspergillus welwitschiae*, although the three isolates present a similar pattern of CAZYmes. CAZymes are responsible for the degradation, synthesis or modification of oligo- and polysaccharides and glycoconjugates. Moreover, many CAZymes play a central role in the breakdown of the plant cell wall (the so-called cell wall degrading enzymes, CWDEs) as well as in host-pathogen interactions (BENOIT et al., 2015). Fungi produce are exceptional producers of CAZymes, sometimes applied in the industry (PARANTHAMAN; ALAGUSUNDARAM; INDHUMATHI, 2009), especially the CWDEs since carbohydrates released from plant cell wall are crucial for the nutrition of both saprotrophic and pathotrophic fungi (LO PRESTI et al., 2015; RODRIGUEZ-MORENO et al., 2018). For plant pathogenic fungi, CWDEs are extremely important for penetration and successful infection of their hosts since localized degradation of the cell wall is necessary for accessing plant cytoplasm as well as spreading across host tissues. Furthermore, many CWDEs, such as pectate lyases, are also related to pathogenicity as mentioned earlies in our discussion (FU et al., 2015; YAKOBY et al., 2007). Considering the main composition of the plant cell wall (celluloses, hemicelluloses, lignins, and pectins), the enzymes cellulases, hemicellulases, ligninases and pectinases are the major plant cell wall degrading

enzymes in phytopathogenic fungi (KUBICEK; STARR; GLASS, 2014). Furthermore, cutinases are also important since they are frequently produced in the early stages of the infection of plants by fungi to degrade the plant cuticle and, thus are important pathogenicity factors in some fungi (OHKURA; COTTY; ORBACH, 2018).

Pectin can be broken down by pectin lyase, pectate lyase, pectin esterase, and polygalacturonase (TAYI et al., 2016). These enzymes mainly fall into nine CAZyme families, including GH28, GH78, GH88, CE8, PL1, PL2, PL3, PL9, and PL10. The three *Aspergillus welwitschiae* isolates displayed a similar profile of genes related to pectin degrading enzymes, and five out of the main nine CAZYmes families were detected in all of them: GH28, GH78, GH88, CE8, and PL1. Nonetheless, minor differences were observed between sisal isolates and that from W. mirabilis for polygalacturonases. Polygalacturonases (GH28 family) play a critical role in pectin degradation in fungal pathogens (LIU et al., 2017). GH28 was one of the CAZY families with the highest relative abundance in sisal isolates (4th and 3rd most abundant in CCMB663 and CCMB674, respectively) in comparison to the *Welwitschia mirabilis* isolate (the seventh most abundant). Many fungi, such as the necrotrophic Sclerotinia sclerotiorum has an expanded family of polygalacturonases (ZHAO et al., 2013), suggesting that our sisal isolates have a high capacity of pectin degradation. Furthermore, the GH78 family, which is also related to pectin degradation, was also quite relatively abundant in all *Aspergillus welwitschiae* isolates, especially in those from sisal plants, suggesting that they have a diversified repertoire for pectin degradation.

Lignocellulose is a complex formed by celluloses, hemicelluloses, and lignins, and comprises the most abundant plant biomass on the planet. Lignocellulose degradation is a complex process that encompasses the cooperation of heterogeneous groups of enzymes, in which most of the lignocellulose degrading enzymes are in glycosyl hydrolases (GH) class (POLLEGIONI; TONIN; ROSINI, 2015), such as cellulases in families GH1, GH3, GH5, GH45, and GH74 (LU et al., 2016), hemicelluloses in families GH3, GH10, GH11, GH51 and GH54 (RAVALASON et al., 2012), and ligninases in families AA1 and AA2. All these families occurred in all *A. welwitschiae* isolates, except for GH45 and GH74 that was not detected in CCMB674 isolate. Thus, they are fully able to degrade the lignocellulose of plant cell walls.

The very high relative abundance of the GH3 family in all *Aspergillus welwitschiae* isolates is remarkable. According to the CAZy database, this family comprises exo-acting β-D-glucosidases, α-L-arabinofuranosidases, β-D-xylopyranosidases, N-acetyl-β-D-glucosaminidases (glycoside hydrolases), and N-acetyl-β-D-glucosaminide phosphorylases. GH3 family was one of the fifth most relatively abundant in all *A. welwitschiae* isolates accounting for 2.6-3.5% of all predicted CAZYmes detected (Figure 13). According to Zhao et al. (2013), two ascomycotan necrotrophic fungi, *Nectria haematococca* and *Fusarium oxysporum* have more predicted GH3 enzymes than any other fungi. Moreover, the GH32 family, which degrades fructose-containing polysaccharides (fructans) that are abundant in sisal stem (the main plant site of fungal infection) (DUARTE et al., 2018), were also quite relatively abundant in all *A. welwitschiae* isolates, which means that the fungi are fully equipped to degrade and feed on the polyssaccharides of the sisal stem.

Cutin is composed of hydroxy and hydroxyepoxy fatty acids, and CE5 family, which comprises cutinases, catalyze the cleavage of cutin ester bonds of cutin to release cutin monomers (FAN, 2002). CE5 family is much more relatively abundant in *Aspergillus welwitschiae* sisal isolates than in CBS 139.54 (Figure 13). This fact is in complete accordance with the structural biology of the sisal plant, whose one of the most prominent morphological features is its thick cuticle layer in both leaf sides (DA CUNHA NETO; MARTINS, 2012). Most necrotrophic plant pathogenic fungi displayed many cutinase coding genes (ZHAO et al., 2013).

## 7 CONCLUSIONS

- We provided two new nuclear genomes for the fungal species *Aspergillus welwitschiae*, which in comparison with the only genome available so far, possesses a smaller gene content, but a greater amount of secreted proteins.
- We provided two new assembled and annotated mitochondrial genomes for the isolates of sisal, that are essentially similar to that of CBS 139.54.
- The secreted proteins of all three isolates contain effector proteins, cell wall-degrading enzymes and other minor protein groups.
- All three isolates possess identifiable effector proteins that are know to be important for virulence in other species.
- All three isolates possess a strong arsenal of carbohydrate active enzymes, including proteins related to the degradation of polysaccharides of the bole of sisal.

## 8 REFERENCES


AKAGI, Y. et al. Horizontal Chromosome Transfer, a Mechanism for the Evolution and Differentiation of a Plant-Pathogenic Fungus. **Eukaryotic Cell**, v. 8, n. 11, p. 1732–1738, 2009.

ALVAREZ-PEREZ, S. et al. Mating type and invasiveness are significantly associated in *Aspergillus fumigatus*. **Medical Mycology**, v. 48, n. 2, p. 273–277, 2010.

ALVES, J. J. A.; ARAÚJO, M. A. DE; NASCIMENTO, S. S. DO. DEGRADAÇÃO DA CAATINGA: UMA INVESTIGAÇÃO ECOGEOGRÁFICA. **Revista Caatinga**, v. 22, n. 3, p. 126–135, 2009.

ANDREWS, S. **FastQC: a quality control tool for high throughput sequence data**Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, , 2010.

ARAUJO, F. A. et al. GO FEAT: A rapid web-based functional annotation tool for genomic and transcriptomic data. **Scientific Reports**, v. 8, n. 1, p. 8–11, 2018.

ASHWIN, N. M. R. et al. CfPDIP1, a novel secreted protein of *Colletotrichum falcatum*, elicits defense responses in sugarcane and triggers hypersensitive response in tobacco. **Applied Microbiology and Biotechnology**, v. 102, n. 14, p. 6001–6021, 2018.

BANKEVICH, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **Journal of computational biology : a journal of computational molecular cell biology**, v. 19, n. 5, p. 455–77, 2012.

BARBOSA, C. S. et al. Genome sequence and effectorome of *Moniliophthora perniciosa* and *Moniliophthora roreri* subpopulations. **BMC Genomics**, v. 19, n. 1, p. 509, 2018.

BECK, N.; LANG, B. MFannot, organelle genome annotation websever. **Canada: Université de Montréal QC**, 2010.

BECK, N.; LANG, B. F. RNAweasel, a webserver for identification of mitochondrial, structured RNAs. **Montreal (Quebec): University of Montreal**, 2009.

BENNETT, J. W.; KLICH, M. Mycotoxins. **Clinical microbiology reviews**, v. 16, n. 3, p. 497–516, 2003.

BENOIT, I. et al. Closely related fungi employ diverse enzymatic strategies to degrade plant biomass. **Biotechnology for Biofuels**, v. 8, n. 1, p. 1–14, 2015.

BENSON, D. A. et al. GenBank. **Nucleic Acids Research**, v. 45, n. D1, p. D37–D42, 2017.

BERNT, M. et al. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. **Molecular Phylogenetics and Evolution**, v. 69, n. 2, p. 313–319, 2013.

BOTURA, M. B. et al. *In vivo* anthelmintic activity of an aqueous extract from sisal waste (*Agave sisalana* Perr.) against gastrointestinal nematodes in goats. **Veterinary Parasitology**, v. 177, n. 1–2, p. 104–110, 2011.

BRAAKSMA, M. et al. An inventory of the *Aspergillus niger* secretome by combining *in silico* predictions with shotgun proteomics data. **BMC Genomics**, v. 11, n. 1, p. 1–11, 2010.

BUSHNELL, B. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. **Joint Genome Institute. https://jgi. doe. gov/data-and-tools/bbtools**, 2018.

CAIRNS, T. C.; NAI, C.; MEYER, V. How a fungus shapes biotechnology: 100 years of *Aspergillus niger* research. **Fungal Biology and Biotechnology**, v. 5, n. 1, p. 1–14, 2018.

CAMACHO, C. et al. BLAST+: architecture and applications. **BMC bioinformatics**, v. 10, p.

421, 2009.

CHAUDHARY, N.; MARR, K. A. Impact of *Aspergillus fumigatus* in allergic airway diseases. **Clinical and Translational Allergy**, v. 1, n. 1, p. 1–7, 2011.

CHELLAPANDI, P. Production and preliminary characterization of alkaline protease from *Aspergillus flavus* and *Aspergillus terreus*. **E-Journal of Chemistry**, v. 7, n. 2, p. 479–482, 2010.

CUESTA-ASTROZ, Y. et al. Helminth secretomes reflect different lifestyles and parasitized hosts. **International Journal for Parasitology**, v. 47, n. 9, p. 529–544, 2017.

DA CUNHA NETO, I. L.; MARTINS, F. M. Anatomy of the vegetative organs of *Agave sisalana* Perrine ex engelm (Agavaceae). **Revista Caatinga**, v. 25, n. 2, 2012.

DARLING, A. C. E. et al. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. **Genome Research**, v. 14, p. 1394–1403, 2004.

DEAN, R. et al. The Top 10 fungal pathogens in molecular plant pathology. **Molecular plant pathology**, v. 13, n. 4, p. 414–30, 2012.

DEBNATH, M. et al. Biotechnological intervention of *Agave sisalana*: a unique fiber yielding plant with medicinal property. **J. Med. Plants Res,** v. 4, n. 3, p. 177–187, 2010.

DIERCKXSENS, N.; MARDULYN, P.; SMITS, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. **Nucleic Acids Research**, v. 45, n. 4, 2017.

DODDS, P. N.; RATHJEN, J. P. Plant immunity: Towards an integrated view of plant-pathogen interactions. **Nature Reviews Genetics**, v. 11, n. 8, p. 539–548, 2010.

DOEHLEMANN, G. et al. Plant Pathogenic Fungi. **Mycrobiology Spectrum**, v. 79, n. 6, p. 919, 2007.

DOS SANTOS, W. N. L. et al. Biosorption of Pb(II) and Cd(II) ions by *Agave sisalana* (sisal fiber). **Microchemical Journal**, v. 97, n. 2, p. 269–273, 2011.

DUARTE, E. A. et al. Putting the mess in order: *Aspergillus welwitschiae* (and not *A. niger*) is the etiologic agent of the sisal bole rot disease. **Frontiers in Microbiology**, v. 9, n. 1227, 2018.

EL-SHEEKH, M. M. et al. Effective technological pectinases by *Aspergillus carneus* NRC1 utilizing the Egyptian orange juice industry scraps. **International Biodeterioration and Biodegradation**, v. 63, n. 1, p. 12–18, 2009.

EL KHOURY, A. E.; ATOUI, A. Ochratoxin a: General overview and actual molecular status. **Toxins**, v. 2, n. 4, p. 461–493, 2010.

EMANUELSSON, O. et al. Locating proteins in the cell using TargetP , SignalP and related tools. **Nature Protocols**, v. 2, n. 4, p. 953–971, 2007.

ESSER, K. **The Mycota - A Comprehensive Treatise on fungi as Experimental Systems for Basic and Applied Research**. [s.l.] Springer US, 2001.

EXPERT, D. et al. *Dickeya dadantii* pectic enzymes necessary for virulence are also responsible for activation of the *Arabidopsis thaliana* innate immune system. **Molecular Plant Pathology**, v. 19, n. 2, p. 313–327, 2018.

FALAHATI, M. et al. Fractionation and identification of the allergic proteins in Aspergillus species. **Current medical mycology**, v. 2, n. 4, p. 37–45, 2016.

FAN, C. Diversity of cutinases from plant pathogenic fungi: differential and sequential expression of cutinolytic esterases by *Alternaria brassicicola*. **FEMS Microbiology Letters**, v. 158, n. 1, p. 33–38, 2002.

FÉLIX, C. et al. Temperature Modulates the Secretome of the Phytopathogenic Fungus *Lasiodiplodia theobromae*. **Frontiers in Plant Science**, v. 7, n. August, p. 1–12, 2016.

FERRANTI, L. S. et al. Occurrence and fumonisin B2 producing potential of Aspergillus section Nigri in Brazil nuts. **Mycotoxin Research**, v. 33, n. 1, p. 49–58, 2017.

FU, L. et al. Characterization of Cell-Death-Inducing Members of the Pectate Lyase Gene Family in *Phytophthora capsici* and Their Contributions to Infection of Pepper . **Molecular Plant-Microbe Interactions**, v. 28, n. 7, p. 766–775, 2015.

GARRIGUES, S.; GANDÍA, M.; MARCOS, J. F. Occurrence and function of fungal antifungal proteins: a case study of the citrus postharvest pathogen *Penicillium digitatum*. **Applied Microbiology and Biotechnology**, v. 100, n. 5, p. 2243–2256, 2016.

GHERBAWY, Y. et al. Molecular Characterization of Black Aspergillus Species from Onion and Their Potential for Ochratoxin A and Fumonisin B2 Production. **Foodborne Pathogens and Disease**, v. 12, n. 5, p. 414–423, 2015.

GIOTI, A. et al. Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. **Molecular Biology and Evolution**, v. 29, n. 10, p. 3215–3226, 2012.

GIRARD, V. et al. Secretomes: The fungal strike force. **Proteomics**, v. 13, n. 3–4, p. 597–608, 2013.

GOFFEAU, A. et al. Life with 6000 Genes. **Science**, v. 274, n. October, p. 546–567, 1996.

GRIGORIEV, I. V. et al. MycoCosm portal: Gearing up for 1000 fungal genomes. **Nucleic Acids Research**, v. 42, n. D1, p. 699–704, 2014.

GUYON, K. et al. Secretome analysis reveals effector candidates associated with broad host range necrotrophy in the fungal plant pathogen *Sclerotinia sclerotiorum*. **BMC Genomics**, v. 15, n. 336, p. 1–18, 2014.

GUZMÁN-GUZMÁN, P. et al. Identification of effector-like proteins in *Trichoderma* spp. and role of a hydrophobin in the plant-fungus interaction and mycoparasitism. **BMC Genetics**, v. 18, n. 16, p. 1–20, 2017.

HEDAYATI, M. T. et al. *Aspergillus flavus*: Human pathogen, allergen and mycotoxin producer. **Microbiology**, v. 153, n. 6, p. 1677–1692, 2007.

JIANG, X. et al. Constitutive expression, purification and characterisation of pectin methylesterase from *Aspergillus niger* in *Pichia pastoris* for potential application in the fruit juice industry. **Journal of the Science of Food and Agriculture**, v. 93, n. 2, p. 375–381, 2013.

KAVANAGH, K. **Fungi - Biology and Applications**. County Kildare, Ireland: John Wiley & Sons, 2011.

KEARSE, M. et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 2012.

KEEGSTRA, K. Plant Cell Walls. **Plant Physiology**, v. 154, p. 483–486, 2010.

KHALDI, N. et al. SMURF: Genomic mapping of fungal secondary metabolite clusters. **Fungal Genetics and Biology**, v. 47, n. 9, p. 736–741, 2010.

KIM, S. et al. MHP1, a *Magnaporthe grisea* hydrophobin gene, is required for fungal development and plant colonization. **Molecular Microbiology**, v. 57, n. 5, p. 1224–1237, 2005.

KIMARO, D. N.; MSANYA, B. M. Review of Sisal Production and Research in Tanzania. **African study Monographs**, v. 15, n. 4, p. 227–242, 1994.

KING, E. D. et al. An industry perspective on the use of “atoxigenic” strains of *Aspergillus flavus* as biological control agents and the significance of cyclopiazonic acid. **Toxin Reviews**, v. 30, n. 2–3, p. 33–41, 2011.

KOECK, M.; HARDHAM, A. R.; DODDS, P. N. The role of effectors of biotrophic and hemibiotrophic fungi in infection. **Cell Microbiol**, v. 13, n. 12, p. 1849–1857, 2013.

KROGH, A. et al. Predicting Transmembrane Protein Topology with a Hidden Markov Model : Application to Complete Genomes. **Journal of Molecular Biology**, v. 305, p. 567–580, 2001.

KUBICEK, C. P.; STARR, T. L.; GLASS, N. L. Plant Cell Wall–Degrading Enzymes and Their Secretion in Plant-Pathogenic Fungi. **Annual Review of Phytopathology**, v. 52, n. 1, p. 427–451, 2014.

KUMAR, S. et al. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. **Briefings in Bioinformatics**, v. 9, n. 4, p. 299–306, 2008.

KUPPIREDDY, V. S. et al. Identification and Initial Characterization of the Effectors of an Anther Smut Fungus and Potential Host Target Proteins. **International Journal of Molecular Sciences**, v. 18, n. 2489, p. 1–19, 2017.

LANGMEAD, B. et al. Scaling read aligners to hundreds of threads on general-purpose processors. **Bioinformatics**, v. 35, n. 3, p. 421–432, 2019.

LÉVEILLARD, T.; AÏT-ALI, N. Cell Signaling with Extracellular Thioredoxin and Thioredoxin-Like Proteins: Insight into Their Mechanisms of Action. **Oxidative Medicine and Cellular Longevity**, v. 2017, p. 1–11, 2017.

LI, Z. et al. Candidate effector proteins of the necrotrophic apple canker pathogen *Valsa mali* can suppress BAX-induced PCD. **Frontiers in Plant Science**, v. 6, n. July, p. 1–9, 2015.

LINDGREEN, S. AdapterRemoval: Easy cleaning of next-generation sequencing reads. **BMC Research Notes**, v. 5, 2012.

LIU, C. Q. et al. Polygalacturonase gene pgxB in *Aspergillus Niger* is a virulence factor in apple fruit. **PLoS ONE**, v. 12, n. 3, p. 43, 2017.

LO PRESTI, L. et al. Fungal Effectors and Plant Susceptibility. **Annual Review of Plant Biology**, v. 66, n. 1, p. 513–545, 2015.

LORANG, J. Necrotrophic Exploitation and Subversion of Plant Defense: A Lifestyle or Just a Phase, and Implications in Breeding Resistance. **Phytopathology**, v. 109, n. 3, p. 332–346, 2018.

LU, X. et al. Adsorption and mechanism of cellulase enzymes onto lignin isolated from corn stover pretreated with liquid hot water. **Biotechnology for Biofuels**, v. 9, n. 1, p. 1–12, 2016.

MACHADO, E. M. S. et al. Growth of fungal strains on coffee industry residues with removal of polyphenolic compounds. **Biochemical Engineering Journal**, v. 60, p. 87–90, 2012.

MANSFIELD, J. et al. Top 10 plant pathogenic bacteria in molecular plant pathology. **Molecular Plant Pathology**, v. 13, n. 6, p. 614–629, 2012.

MASSI, F. P. et al. Prospecting for the incidence of genes involved in ochratoxin and fumonisin biosynthesis in Brazilian strains of *Aspergillus niger* and *Aspergillus welwitschiae.* **International Journal of Food Microbiology**, v. 221, p. 19–28, 2016.

MEDEMA, M. H. et al. antiSMASH : rapid identification , annotation and analysis of secondary

metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **Nucleic Acids Research**, v. 39, n. June, p. 339–346, 2011.

MELÉNDEZ, H. G. et al. Role of the *Botrytis cinerea* FKBP12 ortholog in pathogenic development and in sulfur regulation. **Fungal Genetics and Biology**, v. 46, n. 4, p. 308–320, 2009.

MÉSZÁROS, B.; ERDÖS, G.; DOSZTÁNYI, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. **Nucleic Acids Research**, v. 46, n. W1, p. W329–W337, 2018.

MHAIKI, C. J. N. et al. ISOLATION AND CHARACTERISATION OF MICRO-ORGANISMS WITH INDUSTRIAL IMPORTANCE FROM SISAL BOLE ROTS. **Tanzania Journal of Engineering and Technology**, v. 34, n. 1, 2013.

MIRANDA, F. et al. Improving metagenomic assemblies through data partitioning: A GC content approach. **bioRxiv**, p. 415–425, 2018.

MOONJELY, S.; KEYHANI, N. O.; BIDOCHKA, M. J. Hydrophobins contribute to root colonization and stress responses in the rhizosphere-competent insect pathogenic fungus Beauveria bassiana. **Microbiology (United Kingdom)**, v. 164, n. 4, p. 517–528, 2018.

MOORE, D.; ROBSON, G. D.; TRINCI, A. P. J. **21st Century Guidebook to Fungi**. Cambridge, UK: [s.n.].

MORETTI, A.; SUSCA, A. **Mycotoxigenic Fungi**. [s.l.] Humana Press, 2017. v. 1542

NAGANAGOUDA, K.; SALIMATH, P. V.; MULIMANI, V. H. Purification and characterization of endo-β-1,4 mannanase from *Aspergillus niger* gr for application in food processing industry. **Journal of Microbiology and Biotechnology**, v. 19, n. 10, p. 1184–1190, 2009.

NITSCHE, B. M. et al. The carbon starvation response of *Aspergillus niger* during submerged cultivation: Insights from the transcriptome and secretome. **BMC Genomics**, v. 13, n. 1, 2012.

OHKURA, M.; COTTY, P. J.; ORBACH, M. J. Comparative Genomics of *Aspergillus flavus* S and L Morphotypes Yield Insights into Niche Adaptation. **G3: Genes | Genomes | Genetics**, v. 8, n. December, p. 3915–3930, 2018.

PALLAS, V.; GARCÍA, J. A. How do plant viruses induce disease? Interactions and interference with host components. **Journal of General Virology**, v. 92, n. 12, p. 2691–2705, 2011.

PALUMBO, J. D.; O'KEEFFE, T. L. Detection and discrimination of four Aspergillus section Nigri species by PCR. **Letters in Applied Microbiology**, v. 60, n. 2, p. 188–195, 2015.

PARANTHAMAN, R.; ALAGUSUNDARAM, K.; INDHUMATHI, J. Production of protease from rice mill wastes by *Aspergillus niger* in solid state fermentation. **World Journal of Agricultural Sciences**, v. 5, n. 3, p. 308–312, 2009.

PARIS, S. et al. Conidial Hydrophobins of Aspergillus fumigatus. **APPLIED AND ENVIRONMENTAL MICROBIOLOGY**, v. 69, n. 3, p. 1581–1588, 2003.

PETERSEN, T. N. et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature Methods**, v. 8, n. 10, p. 785–786, 2011.

PETTOLINO, F. A. et al. Determining the polysaccharide composition of plant cell walls. **Nature Protocols**, v. 7, n. 9, p. 1590–1607, 2012.

POLLEGIONI, L.; TONIN, F.; ROSINI, E. Lignin-degrading enzymes. **FEBS Journal**, v. 282, n. 7, p. 1190–1213, 2015.

R CORE TEAM. R: A language and environment for statistical computing. 2013.

RAVALASON, H. et al. *Fusarium verticillioides* secretome as a source of auxiliary enzymes to enhance saccharification of wheat straw. **Bioresource Technology**, v. 114, p. 589–596, 2012.

RODRIGUEZ-MORENO, L. et al. Tools of the crook- infection strategies of fungal plant pathogens. **Plant Journal**, v. 93, n. 4, p. 664–674, 2018.

ROSE, L. J. et al. Preharvest Management Strategies and Their Impact on Mycotoxigenic Fungi and Associated Mycotoxins. **Intech open**, v. 2, p. 64, 2014.

RYTIOJA, J. et al. Plant-Polysaccharide-Degrading Enzymes from Basidiomycetes. **Microbiology and Molecular Biology Reviews**, v. 78, n. 4, p. 614–649, 2014.

SANTOS, P. O. DOS et al. Additional species of Aspergillus causing bole rot disease in *Agave sisalana*. **Tropical Plant Pathology**, v. 39, n. 4, p. 331–334, 2014.

SANTOS, J. D. G. et al. Antimicrobial activity of *Agave sisalana*. **African Journal of Biotechnology**, v. 8, n. 22, p. 6181–6184, 2009.

SILVEIRA, R. X. et al. Action of sisal (*Agave sisalana*, Perrine) extract in the *in vitro* development of sheep and goat gastrointestinal nematodes. **Experimental Parasitology**, v. 131, n. 2, p. 162–168, 2012.

SIMÃO, F. A. et al. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. **Bioinformatics**, v. 31, n. 19, p. 3210–3212, 2015.

SONAH, H.; DESHMUKH, R. K.; BÉLANGER, R. R. Computational Prediction of Effector Proteins in Fungi: Opportunities and Challenges. **Frontiers in Plant Science**, v. 7, n. February, p. 1–14, 2016.

SPERSCHNEIDER, J. et al. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. **Molecular Plant Pathology**, v. 19, n. 9, p. 2094–2110, 2018.

STANKE, M. et al. AUGUSTUS: A web server for gene finding in eukaryotes. **Nucleic Acids Research**, v. 32, n. WEB SERVER ISS., p. 309–312, 2004.

STRUCK, C. Infection strategies of plant parasitic fungi. **The Epidemiology of Plant Diseases**, p. 117–137, 2006.

SUSCA, A. et al. Variation in the fumonisin biosynthetic gene cluster in fumonisin-producing and nonproducing black aspergilli. **Fungal Genetics and Biology**, v. 73, p. 39–52, 2014.

SUSCA, A. et al. Variation in fumonisin and ochratoxin production associated with differences in biosynthetic gene content in *Aspergillus niger* and *A. welwitschiae* isolates from multiple crop and geographic origins. **Frontiers in Microbiology**, v. 7, n. SEP, 2016.

TAYI, L. et al. Identification of pectin degrading enzymes secreted by *Xanthomonas oryzae* pv. Oryzae and determination of their role in virulence on rice. **PLoS ONE**, v. 11, n. 12, p. 1–15, 2016.

THORPE, P. et al. Shared transcriptional control and disparate gain and loss of aphid parasitism genes. **Genome Biology and Evolution**, v. 10, n. 10, p. 2716–2733, 2018.

URBAN, M. et al. PHI-base: A new interface and further additions for the multi-species pathogen-host interactions database. **Nucleic Acids Research**, v. 45, n. D1, p. D604–D610, 2017.

VAN DEN BRINK, J.; DE VRIES, R. P. Fungal enzyme sets for plant polysaccharide degradation. **Applied Microbiology and Biotechnology**, v. 91, n. 6, p. 1477–1492, 2011.

VARGA, J. et al. Occurrence of black Aspergilli in indoor environments of six countries. **Arhiv**

**za Higijenu Rada i Toksikologiju**, v. 65, n. 2, p. 219–223, 2014.

VENTER et al. The sequence of the human genome. **Science**, v. 291, n. February, p. 1304–1351, 2001.

VESTH, T. C. et al. Investigation of inter- and intraspecies variation through genome sequencing of Aspergillus section Nigri. **Nature Genetics**, v. 50, n. December, p. 1688–1695, 2018.

VRIES, R. P. DE; GRIGORIEV, I. V.; TSANG, A. **Fungal Genomics**. [s.l: s.n.]. v. 722

WANG, X. et al. The role of effectors and host immunity in plant–necrotrophic fungal interactions. **Virulence**, v. 5, n. 7, p. 722–732, 2014.

WARTENBERG, D. et al. Secretome analysis of *Aspergillus fumigatus* reveals Asp-hemolysin as a major secreted protein. **International Journal of Medical Microbiology**, v. 301, n. 7, p. 602–611, 2011.

YAKOBY, N. et al. *Colletotrichum gloeosporioides* pelB Is an Important Virulence Factor in Avocado Fruit-Fungus Interaction . **Molecular Plant-Microbe Interactions**, v. 14, n. 8, p. 988–995, 2007.

YANDELL, M.; HOLT, C. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. **BMC Bioinformatics**, v. 12, n. 1, p. 491, 2011.

ZHANG, H. et al. DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. **Nucleic Acids Research**, v. 46, n. W1, p. W95–W101, 2018.

ZHAO, Z. et al. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. **BMC Genomics**, v. 14, n. 274, 2013.

ZINJARDE, S. S.; PANT, A. Crude oil degradation by free and immobilized cells of Yarrowia lipolytica NCIM 3589. **Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering**, v. 35, n. 5, p. 755–763, 2000.

ZVEREVA, A. S.; POOGGIN, M. M. Silencing and innate immunity in plant defense against viral and non-viral pathogens. **Viruses**, v. 4, n. 11, p. 2578–2597, 2012.

# 9 Supplementary Material

## Pipeline dependencies

This pipeline and data sets were processed in the sagarana HPC cluster, CEPAD-ICB-UFMG and in the LGCM/Aquacen server. The following programs, scripts and databases were used:

## Programs

1 FastQC v0.11.5 (SIMON, 2010)
(https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)
2 AdapterRemoval v2 (LINDGREEN, 2012)
(https://github.com/MikkelSchubert/adapterremoval)
3 GCSplit (MIRANDA et al., 2018) (https://github.com/mirand863/gcsplit)
4 SPAdes (BANKEVICH et al., 2012)
5 MAKER2 v2.31.9 (YANDELL; HOLT, 2011)
6 BUSCO (SIMÃO et al., 2015)
7 BBDuk (BUSHNEL, 2014)
8 Bowtie2 (LANGMEAD et al., 2019)
9 BLASTn (CAMACHO et al., 2009)
10 MITOS2 (BERNT et al., 2013)
11 RNAweasel (BECK & LANG, 2009)
12 MFannot (BECK & LANG, 2010)
13 geneious (KEARSE et al., 2012)
14 SMURF (KHALDI et al., 2010)
15 antiSMASH (MEDEMA et al., 2011)
16 SignalP 4.1 (PETERSEN et al., 2011)
17 TargetP (EMANUELSSON et al., 2007)
18 Go Feat (ARAUJO et al., 2018)

## Scripts

1 *scaffold_stats.pl* (https://github.com/blaxterlab/scripts/blob/master/tools/scaffold_stats.pl)
2 *count.py*

```
c = open("<OUTFILE>", "w")
with open("<INFILE>.txt", "r") as f:
    lista = [line.strip() for line in f]
for x in b:
    print(x, "\t", lista.count(x), file=c)
```

3 *heatmap.R*

```
abund_0 <- read.delim("<PATH>.txt")
abund_2$CCMB663<-abund_2$CCMB663/461*100
abund_2$CCMB674<-abund_2$CCMB674/612*100
abund_2$CBS.139.54<-abund_2$CBS.139.54/429*100
library("RColorBrewer")
pdf(file = "<OUTFILE>.pdf", width = 10, height =15)
heatmap(as.matrix(abund_2),scale="column",col=colorRampPalette(
brewer.pal(8, "Blues"))(25),main="GO Terms
Distribution",Rowv=NA,Colv=NA)
dev.off()
```

## Databases

1 JGI – Mycocosm (GRIGORIEV et al., 2014)
2 NCBI – GenBank (BENSON et al., 2017)
3 PHI-base – PHIB-blast (URBAN et al., 2017)

4 CAZyme (ZHANG et al., 2018)

The pipeline consisted of the following steps:

1 Quality check

Raw .fastq sequences quality was accessed with FastQC.

```
fastqc <FASTQ file>
```

2 Sequences trimming and filtering

Low-quality bases (Phred < 20) and adapters were trimmed and overlapping sequences were collapsed with AdapterRemoval

```
AdapterRemoval --file1 <Forward FASTQ file> --file2 <Reverse FASTQ file> --basename PE150 --trimns --trimqualities –collapse
```

3 FASTQ file partitioning

Files were split into four sets with similar GC content with GCSplit.

```
./gcsplit -f arq1.fastq -r arq2.fastq -p 4 -o saida -t 12
```

4 Assembly *per se*

Paired end FASTQ files were assembled with SPAdes.

```
spades.py --pe1-1 <Forward FASTQ file> --pe1-2 <Reverse FASTQ file> --careful --cov-cutoff off -t 12 -o <output PATH>
```

5 Genome annotation

Assembled genomes were annotated with MAKER2.

```
maker <FASTA file>
```

*ab initio prediction*

```
gff3_merge -d ../pyu_contig1.maker.output/pyu_contig1_master_datastore_index.log
maker2gff pyu_contig1.all.gff
fathom -categorize 1000 genome.ann genome.dna
fathom -export 1000 -plus uni.ann uni.dna
forge export.ann export.dna
hmm-assembler.pl snap . > ../snap1.hmm
maker -base snap1
```

6 Assembly QC (Quali-quantitative)

Expected gene content was accessed with BUSCO. The default reference for ascomycetes is *Aspergillus nidulans*, the closest available species.

```
export BUSCO_CONFIG_FILE="/home/pfonseca/databases/busco_dataset/config.ini"
export PATH="/opt/augustus/Augustus/bin:$PATH"
export PATH="/opt/augustus/Augustus/scripts:$PATH"
export AUGUSTUS_CONFIG_PATH="/home/pfonseca/databases/config/"
run_BUSCO.py -i <FASTA file> -o <run name> -l ascomycota_odb9 -m geno
```

7 Assembly QC (Quantitative)

Assembly statistics were accessed with *scaffold_stats.pl.*

***perl scaffold_stats.pl -c 0 -f <FASTA file>***

8 Sequences trimming and filtering (Mitochondria)

Low-quality regions (Phred < 20) and adapter sequences were removed with BBDuk.

***bbduk.sh in1=<Forward FASTQ file> in2=<Reverse FASTQ file> out1=<output Forward FASTQ file> out2=<output Reverse FASTQ file> ref=<PATH to adapters.fa> ktrim=r k=23 mink=11 hdist=1 tpe tbo***

9 Mapping sequences to reference

Reads in paired end FASTQ files were mapped to a reference genome (*Aspergillus welwitschiae* CBS 139.54) mitochondrial genome with Bowtie2.

**/usr/bin/bowtie2-build <PATH_TO_FASTA> <OUT_LIBRARY>**
**/usr/bin/bowtie2 -x <OUT_LIBRARY> -1 <PATH_TO_FASTQ>.fq -2 <PATH_TO_FASTA>.fq --al-conc <OUTFILE>.fq -S <OUTFILE>.sam**

10 Comparison against a Mitochondrial database

Comparison against Mitochondrial RefSeq was performed with BLASTn.

***blastn -query <assembled FASTA file> -db refseqmit.fna -num_threads 12 -perc_identity 80 -outfmt 7 -out <output TXT file>***

**Table S8:** Distribution of GO terms on CCMB663, CCMB674 and CBS 139.54

| Condition | Gene Ontology number (GO:) |
|---|---|
| **Shared by CCMB663, CCMB674 and CBS 139.54 (63 terms)** | 0008233 0004650 0004672 0061634 0008152 0003824 0008422 0005199 0005524 0005506 0016705 0004553 0047911 0005975 0005886 0004190 0008081 0005509 0047490 0009277 0016740 0003723 0016829 0016021 0071555 0016798 0016788 0030246 0004067 0016020 0005623 0045493 0006629 0004222 0008237 0008061 0005576 0016491 0071949 0008810 0052692 0003676 0004185 0019031 0006528 0008168 0016787 0102483 0020037 0045490 0050525 0004180 0046558 0003993 0031225 0004177 0000272 0008270 0004252 0005634 0031222 0046872 0030245 |
| **Shared by CCMB663 and CBS 139.54 (4 terms)** | 0033946 0045454 0051213 0005615 |
| **Shared by CCMB674 and CCMB663 (58 terms)** | 0006364 0004623 0015926 0004563 0000742 0003677 0006121 0004348 0046555 0042546 0031047 0006078 0016813 0050660 0030599 0008236 0005507 0018293 0016614 0102148 0004540 0016042 0048288 0003979 0004181 0004565 0006644 2000899 0016805 0016158 0016791 0004339 0004806 0031218 0009395 0046556 0004622 0006979 0006665 0005759 0006030 0046373 0006397 0102210 0006351 0005618 0004601 0016977 0016162 0004503 0000981 0050482 0052716 0042545 2001070 0016691 0045330 0051287 |
| **Shared by CBS 139.54 and CCMB674 (6 terms)** | 0031176 0016985 0008783 0045735 0016052 0016853 |
| **Exclusive to CCMB663 (16 terms)** | 0006801 0016872 0047682 0043864 0047739 0030570 0031408 0004040 0000155 0019748 0052761 0008239 0000009 0047646 0006508 0000943 |
| **Exclusive to CBS 139.54 (28 terms)** | 0004556 0004521 0003756 0006458 0000287 0004674 0006164 0103025 0006080 0005788 0003861 0016868 0004643 0030127 0034314 0006457 0005885 0005789 0003937 0051082 0009098 0051539 0006270 0009316 0006886 0000139 0006281 0006888 |
| **Exclusive to CCMB674 (53 terms)** | 0008496 0009116 0003735 0019863 0102131 0008199 0033897 0004497 0005840 0006529 0004725 0043333 0030328 0042744 0031314 0006744 0006414 0004316 0004096 0000225 0016407 0000324 0007062 0006308 0034480 0016702 0030596 0043324 0045132 0009251 0001735 0042578 0042803 0005773 0070940 0004459 0006506 0006725 0102132 0032366 0033609 0004359 0004519 0030600 0008420 0030248 0000070 0016311 0006890 0019752 0004322 0046564 0004066 |

**Figure S14:** Graphical representation for the *Aspergillus welwitschiae* CCMB674 mitochondrial genome.

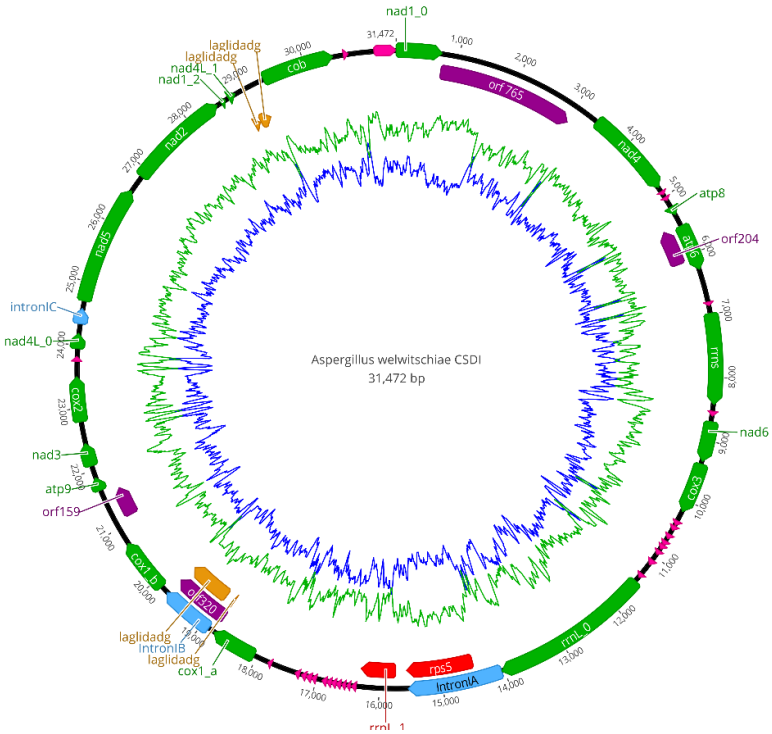



**Figure S15:** Graphical representation for the *Aspergillus welwitschiae* CCMB663 mitochondrial genome.