



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

RENATO AUGUSTO CORRÊA DOS SANTOS

ANÁLISE GENÔMICA DA LEVEDURA XILANOLÍTICA
Pseudozyma brasiliensis GHG001

CAMPINAS

2018

RENATO AUGUSTO CORRÊA DOS SANTOS

**ANÁLISE GENÔMICA DA LEVEDURA XILANOLÍTICA
Pseudozyma brasiliensis GHG001**

Dissertação apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do Título de Mestre em Genética e Biologia Molecular na área de Bioinformática.

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO RENATO AUGUSTO CORRÊA DOS SANTOS E ORIENTADA PELO DR. GUSTAVO HENRIQUE GOLDMAN.

Orientador: Prof. Dr. Gustavo Henrique Goldman

Co-Orientador: Dr. rer. nat. Diego Mauricio Riaño Pachón

CAMPINAS

2018

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2014/15799-7; FAPESP, 2016/13968-1
ORCID: <https://orcid.org/0000-0003-0826-5479>

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

Sa59a Santos, Renato Augusto Corrêa dos, 1990-
Análise genômica da levedura xilanolítica *Pseudozyma brasiliensis*
GHG001 / Renato Augusto Corrêa dos Santos. – Campinas, SP : [s.n.], 2018.

Orientador: Gustavo Henrique Goldman.
Coorientador: Diego Mauricio Riaño Pachón.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Genômica. 2. Comparação genômica. 3. RNA-seq. 4. *Ustilaginaceae*. 5. Pentoses - Metabolismo. I. Goldman, Gustavo Henrique. II. Riaño-Pachón, Diego Mauricio. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Genomic analysis of the xylanolytic yeast *Pseudozyma brasiliensis*
GHG001

Palavras-chave em inglês:

Genomics
Comparative genomics
RNA-seq
Ustilaginaceae
Pentoses - Metabolism

Área de concentração: Bioinformática

Titulação: Mestre em Genética e Biologia Molecular

Banca examinadora:

Gustavo Henrique Goldman [Orientador]
Gabriela Felix Persinoti

Rafael Silva Rocha

Data de defesa: 08-02-2018

Programa de Pós-Graduação: Genética e Biologia Molecular

Campinas, 08 de fevereiro de 2018.

COMISSÃO EXAMINADORA

Prof. Dr. Gustavo Henrique Goldman (orientador)

Dra. Gabriela Felix Persinoti

Prof. Dr. Rafael Silva-Rocha

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

DEDICATÓRIA

**Dedico este trabalho à minha mãe, Alcione Christina Corrêa, meu irmão,
Mauricio Corrêa dos Santos e meu pai, Waldeglace Rodrigues dos Santos,
exemplos de luta e perseverança.**

AGRADECIMENTOS

Diretamente ligados ao meu projeto de mestrado

Ao Dr. Gustavo Henrique Goldman (USP), pela orientação, em particular no conhecimento sobre genética de fungos e pelos desenhos experimentais para as análises de transcriptômica de *Kalmanozyma brasiliensis* GHG001 e pela correção da dissertação.

Ao Dr. rer. nat. Diego Mauricio Riaño Pachón (Universidade de São Paulo, USP; Centro Nacional de Pesquisa em Energia e Materiais, CNPEM), pela orientação nas análises de Bioinformática apresentadas neste trabalho.

À Dra. Juliana Velasco de Castro Oliveira (CNPEM), pelas análises envolvendo as leveduras sequenciadas e de transcriptômica de *K. brasiliensis* GHG001 e pela revisão a dissertação, e aos seus alunos, o Me. Gustavo Pagotto Borin, Me. Thuanny Andrade Borges e Dr. Paulo Vitor Marques Simas, que trabalharam nos experimentos gerados para as análises de transcriptômica apresentadas neste trabalho.

Ao Prof. Dr. Carlos Augusto Rosa (Universidade Federal de Minas Gerais, UFMG), que nos forneceu os isolados de leveduras de *Heliconia psittacorum*, cujos genomas são montados neste trabalho.

Ao Dr. Jason Papin e membros do Papin Lab (Universidade de Virginia, VA, EUA), em especial à aluna de doutorado Maureen Carey e ao Dr. Phillip Yen, que deram valiosas sugestões de curadoria do rascunho do modelo de redes metabólicas de *K. brasiliensis* GHG001.

Ao Dr. Antonis Rokas (Vanderbilt University, TN, USA) e membros do Rokas Lab, em especial à aluna de doutorado Abigail Lind e Dr. Xingxing Shen, que me receberam no laboratório durante o estágio realizado nos Estados Unidos.

Ao aluno de doutorado Christian Lieven e ao Dr. Thiemo Zambanini, que forneceram o modelo metabólico de *Mycosarcoma* (= *Ustilago*) *maydis*, base para o melhoramento do modelo metabólico de *K. brasiliensis* GHG001.

Ao Dr. Renato Vicentini dos Santos, Dr. Marcelo Falsarella Carazzolle e Dra. Gabriela Felix Persinoti (CTBE, CNPEM) pelas valiosas sugestões durante a qualificação do mestrado, contribuindo para a finalização do trabalho; agradeço especialmente à Gabriela pelo suporte computacional durante o período que estive no CTBE/CNPEM.

À Dra. Claudia Barros Monteiro Vitorello (USALQ, USP) pelas valiosas discussões sobre biologia de *Sporisorium scitamineum* e interação patógeno-hospedeiro com cana-de-açúcar, que desde a graduação contribuíram para o entendimento deste grupo de fungos basidiomicetos.

Agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (processos 2014/15799-7 e 2016/13968-1).

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (processo 184722/2014-4).

À FAPESP / Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP), projeto UNICAMP/FINEP-MCT/FAPESP, cujos recursos foram utilizados durante o desenvolvimento deste projeto.

Contribuições significativas para a minha formação como pesquisador

Ao doutorando Lucas de Souza Lopes e aos outros alunos que participaram dos encontros para estudo de Python no CTBE/CNPEM.

Ao Dr. Cameron Mura (University of Virginia, USA) e aos membros do Mura Lab, em especial ao Dr. Charles E. McAnany, pela recepção em seu laboratório e pelas valiosas discussões sobre iniciativas educacionais na área de Bioinformática.

Ao Dr. Ricardo Cerri (Laboratório de Bioinformática e Aprendizado de Máquina, DC, UFSCar), Dr. Carlos Norberto Fischer (UNESP Rio Claro), Alisson H. da Costa, Luiza Leme, Henrique Frajacomo e Leonardo Utida.

À Dra. Ane Hackbart de Medeiros, Dr. Douglas Silva Domingues, Dra. Monalisa Sampaio Carneiro, Dra. Roberta, Dra. Sandra Regina Ceccato Antonini, pelo apoio durante a promoção de encontros na área de Bioinformática aos membros do Grupo de Estudos em Bioinformática e Biologia Molecular no Centro de Ciências Agrárias da UFSCar, em Araras, principalmente alunos graduandos em Biotecnologia.

Ao Raniere Gaia Silva, Leonardo Martins, Franciele Grego Esteves, Ana Ciconelle, Bruno F. Souza, Alisson Hayasi da Costa pelo apoio durante o desenvolvimento do Workshop de Introdução ao Python realizado no Instituto de Química da Universidade de São Paulo.

Ao time do iGEM USP 2017, pelo aprendizado durante as poucas reuniões durante o fim do projeto de mestrado.

À aluna de doutorado Marian Schmidt e ao Dr. Timothy (Tim) James (Universidade de Michigan).

Dra. Flávia Vischi Winck (Laboratório de Biologia de Sistemas Regulatórios, IQ, USP), ao Me. Paulo Alexandrino e à Dra. Aline Silva Romão-Dumaresq pelas discussões valiosas sobre Biologia de Sistemas.

Aos amigos

Maria Julia Magazoni, Mitra Katherina Ferreira, Franciele Grego Esteves, Joeslani Simão Andrade, Luis Felipi Monelli, Gustavo Pagotto Borin, Ana Maria Zetty, Maria Florencia Martinez, Waleska Rojas, Mario Meza, Ledao Chen, Marcio de Souza, Estelita Simões Ribeiro, Ana Silvia de Almeida Scarcella, Raissa Ludwig, Marcela Godoy, Isadora Vitti Vieira, Daiane Belgini, Jessica Carolina Gomes da Silva, Omar Julio Sosa, Natalia de Oliveira Marin, Rafael Fávero (Socorro), Amanda Ribeiro Martins da Silva, Regiane Alves (Socorro), Sibeli De Carli, Ana Eliza, Natalia Gregório, Brenda Dornelles, Leila Melari e Rob, Rose Varella, Itagiba Martins Pereira Filho, Marcos Simões, Antony Graber, Daniel Castro, Jéssica Karine Marques, Caique Carvalho, Renato da Silva Cardoso, Giovana Piovesan, Laura Piovesan, Maria Amélia, e muitos outros personagens essenciais!

À namorada

Letícia Cristina Bento, pelo amor e pela paciência.

RESUMO

Os fungos do Filo Basidiomycota são os organismos mais relevantes na decomposição de matéria orgânica no planeta, sendo que os membros do subfilo Agaricomycotina são as principais espécies envolvidas na desconstrução de biomassa lignocelulósica. A família *Ustilaginaceae* pertence ao subfilo Ustilaginomycotina, grupo-irmão de Agaricomycotina, e inclui importantes fitopatógenos de gramíneas e leveduras não patogênicas que têm se destacado pelo potencial biotecnológico, incluindo a produção de enzimas de degradação de polímeros da parede celular de plantas. O xilano é o principal componente da porção hemicelulósica da parede celular de plantas e o entendimento de sua degradação por fungos, bem como do transporte de seu monômero, a xilose, para o interior celular e seu metabolismo interno, são essenciais para a aplicação industrial e o melhoramento genético destes organismos, incluindo a produção de bioetanol de segunda geração. A levedura xilanolítica *Kalmanozyma brasiliensis* GHG001 (previamente *Pseudozyma brasiliensis*) é um membro de *Ustilaginaceae* isolada do trato intestinal de uma larva de *Chrysomelidae* que parasitava a cana-de-açúcar, durante o screening de leveduras capazes de fermentar xilose, tendo se destacado pela produção de uma endoxilanase da família GH11. As abordagens ômicas são valiosas para a exploração do potencial de degradação de biomassa lignocelulósica. Trabalhos previamente realizados incluem genômica comparativa de diversos grupos de fungos ascomicetos e basidiomicetos com foco na degradação de lignocelulose e a transcriptômica com RNA-seq em basidiomicetos já foi estudada para entendimento da degradação de lignocelulose limitando-se, no entanto, aos membros de Agaricomycotina. Apesar do seu potencial biotecnológico e da disponibilidade das sequências genômicas de diversos membros, a família *Ustilaginaceae* permanece subexplorada. Neste trabalho, abordagens ômicas são usadas para a exploração do potencial biotecnológico desta família de fungos com foco na degradação de xilano e no metabolismo de xilose. Inicialmente, a análise de transcriptômica com RNA-seq é usada para melhorar a compreensão do metabolismo de xilano e xilose em *K. brasiliensis* GHG001, incluindo a expressão de genes que codificam as enzimas de degradação de xilano, os transportadores putativos de xilose e possíveis proteínas reguladoras da transcrição. A análise de RNA-seq foi usada também para o melhoramento da anotação estrutural desta levedura. Em seguida, uma abordagem de genômica comparativa foi aplicada para a

exploração de outros membros da família com genomas sequenciados, incluindo leveduras isoladas de flores de *Heliconia psittacorum*, *Moesziomyces spp.* e *Pseudozyma sp.* pro tempore, potencialmente novas espécies. As análises de RNA-seq melhoraram significativamente a anotação estrutural do genoma de *K. brasiliensis* GHG001 e contribuíram para o melhor entendimento do metabolismo de pentoses e da regulação da expressão gênica. As análises filogenéticas de *Ustilaginaceae* contribuíram para a Sistemática da família de fungos e permitiram não apenas alocar as novas leveduras sequenciadas, mas também possibilitaram o estudo da genômica comparativa. Por fim, a prevalência de Enzimas Ativas em Carboidratos (CAZymes) e a evolução (expansões e contrações) de famílias CAZymes e de Proteínas Associadas à Transcrição (TAPs) destacaram aspectos importantes para o levantamento de hipóteses sobre a família de fungos.

ABSTRACT

Fungi in phylum Basidiomycota comprise the most relevant in decomposition of organic matter in our planet, and members of the Agaricomycotina sub-phylum are the main involved in deconstruction of lignocellulosic biomass. The family *Ustilaginaceae* belongs to sub-phylum Ustilaginomycotina, a sister clade of Agaricomycotina, and includes important grass phytopathogens and non-pathogenic yeasts, both have been highlighted because of their biotechnological potential, including production of enzymes for deconstruction of plant cell wall polymers. Xylan is the main component of the hemicellulosic portion of plant cell wall and the comprehension of its degradation by fungi, as well as the transport of its monomer, xylose, and its internal metabolism are essential for industrial application and genetic improvement of these organisms, including production of Second Generation Bioethanol. The xylanolytic yeast *Kalmanozyma brasiliensis* GHG001 (previously *Pseudozyma brasiliensis*) is a member of *Ustilaginaceae* isolated from the gut of a *Chrysomelidae* larvae that parasitized sugarcane, during a screening of yeasts able to ferment xylose, and was outstanding in overproduction of an endoxylanase of family GH11. Omics approaches are valuable for exploitation of the potential of lignocellulosic biomass deconstruction. Previous works have been carried out that included comparative genomics of several fungal groups within ascomycetes and basidiomycetes with focus on lignocellulose degradation and transcriptomics employing RNA-seq in basidiomycetes has already been carried out for understanding of lignocellulose degradation, but limited to Agaricomycotina. Despite its biotechnological importance and availability of genome sequences for several of its members, the family *Ustilaginaceae* remains underexplored. In this work, omics approaches have been applied to exploit the biotechnological potential of this fungal family with focus on xylan degradation and xylose metabolism. First, a transcriptomic analysis with RNA-seq was employed to improve our comprehension of xylan and xylose metabolism in *K. brasiliensis* GHG001, including the expression of genes coding for xylan degrading enzymes, putative xylose transporters, and possible transcriptional regulatory proteins. RNA-seq analysis was also used to improve the genome structural annotation of the yeast. Next, a comparative genomics approach was used to exploit other members of the fungal family with sequenced genomes, including three novel ustilaginaceous yeasts, possibly new species, isolated from flowers of *Heliconia psittacorum*, *Moesziomyces* spp. F16C1 and F8B2, and

Pseudozyma sp. F5C1 pro tempore, potentially novel species. RNA-seq analyses improved the structural annotation of the *K. brasiliensis* GHG001 genome and allowed a better understanding of the pentose metabolism and regulation of gene expression. Phylogenetic analyses of the family *Ustilaginaceae* contributed to its Systematics, allowing the placement of sequenced yeasts and providing an accurate phylogeny for the comparative genomic analysis. Lastly, the prevalence of CAZymes and the evolution (expansions and contractions) of CAZYme and TAP families highlighted important aspects for raising hypotheses about the fungal family.

LISTA DE ILUSTRAÇÕES

Figura 1 - Ilustração esquemática de um tipo de xilano, o glucuronoarabinoxilano, que apresenta alta quantidade (% peso/ peso, em relação a outros polímeros) na parede celular primária e secundária de gramíneas (imagem modificada de Scheller e Ulvskov, 2010).

Figura 2 - Ação coordenada de enzimas para desconstrução de xilano na produção de biocombustíveis. a - Xilanases, acetil xilano esterases e ácido ferúlico esterases atuam juntas para a geração de pequenos xilo-oligossacarídeos, com a liberação de ácidos ferúlico e acético. b - Arabinofuranosidases e glucuronidases liberam arabinose e ácido glucurônico destes xilo-oligossacarídeos. c - Xilosidases convertem xilo-oligossacarídeos em xilose que os constitui. d - Organismos fermentadores assimilam a xilose e a arabinose, direcionando as pentoses para a Via de Pentoses Fosfato para fermentação de biocombustíveis. Fonte: Dodd e Cann (2009).

Figura 3 - Vias de utilização de xilose em fungos. Fonte: Modificado de Hahn-Hägerdal et al. (2007).

Figura 4 – Fluxo de trabalho de anotação para diferentes tipos de genes. a - Genes codificadores de proteínas em genomas de referência. b - Genes codificadores de proteínas em genomas não considerados referência. c - RNAs não codificadores longos. d - Pequenos RNAs. e - Pseudogenes. Fonte: Mudge e Harrow (2016).

Figura 5 - Diagramas de Venn com genes considerados diferencialmente expressos (DEG) em cada um dos métodos (DESeq2; EdgeR + TMM; Voom + TMM). Em vermelho, o número de genes considerados DE na intersecção entre os métodos considerados, os quais foram usados nas análises subsequentes de enriquecimento de genes e análises de perfil de expressão.

Figura 6 - Genes considerados diferencialmente expressos entre os encontrados exclusivamente na nova anotação (com dados de RNA-seq).

Figura 7 - Análise de perfil de expressão de genes diferencialmente expressos. Genes foram agrupados em seis grupos, de acordo com o perfil de expressão: genes induzidos (XLAIG) e reprimidos (XLARG) em xilano, induzidos (XLOIG) e reprimidos (XLORG) em xilose e genes induzidos (GLUIG) e reprimidos (GLURG) em glicose. O número de genes em cada grupo correspondentes a TAPs, CAZymes e possíveis transportadores de xilose estão destacados em cores vermelho, azul e verde, respectivamente.

Figura 8 - Árvore filogenética Máxima Verossimilhança (ML) de isolados de *Pseudozyma* (antes do grupo polifilético ser dividido em diversos gêneros monofiléticos por Wang et al., 2015) e espécies relacionadas. Para análise filogenética, foram usadas regiões ITS1 e ITS2, 5.8S do rDNA e subunidade D1/D2 da subunidade maior do rDNA. Os valores de bootstrap são baseados em 1000 replicatas.

Figura 9 - Análise filogenética de Máxima Verossimilhança (ML) com supermatriz de aminoácidos e partições para os diferentes marcadores usados por Capella-Gutierrez et al., 2014 com RAxML com 1000 replicatas de bootstrap de Basidiomicetos. Os modelos evolutivos para cada partição foram selecionados de acordo com os critérios da AIC (Critério de Informação Akaike) a partir dos resultados do programa ProtTest3. A família *Ustilaginaceae*, cujos membros são usados nas análises comparativas deste trabalho, está destacada em negrito. *Malassezia globosa* foi usada como grupo externo para as análises filogenética para análises de genômica comparativa, pois era o membro de Ustilaginomycotina fora de *Ustilaginaceae* com genoma sequenciado. Taxa em vermelho correspondem às leveduras com genomas montados neste trabalho.

Figura 10 - Análise filogenética de Máxima Verossimilhança (ML) com supermatriz de aminoácidos e partições para os diferentes marcadores usados por Capella-Gutierrez et al., 2014 com RAxML com 1000 replicatas de bootstrap de *Ustilaginaceae*. Os modelos evolutivos para cada partição foram selecionados de acordo com os critérios da AIC (Critério de Informação Akaike) a partir dos resultados do programa ProtTest3.

Figura 11 - MrBayes v3.2.7 usando os quatro marcadores de Capella-Gutierrez, et al. (2014). A inferência filogenética foi realizada com partições, usando os modelos evolutivos para cada marcador de acordo com a tabela 10, além de taxas de evolução gama, taxas variáveis para cada partição (ratepr=variable), e relógio uniforme. O programa foi rodado usando 100.000 gerações.

Figura 12 - Análise filogenética de Máxima Verossimilhança (ML) com supermatriz de aminoácidos usando o método implementado no FastTree e todos os ortólogos de cópia única + ortólogos de cópia única com um organismo ausente. O modelo evolutivo WAG+G foi aplicado.

Figura 13 - Inferência usando IQ-TREE com todos os ortólogos de cópia única + ortólogos de cópia única com um organismo ausente. O modelo evolutivo WAG+G foi aplicado. Os valores de bootstrap são baseados em 1000 replicatas.

Figura 14 - Análise de presença e ausência de ortólogos de cópia única nos proteomas de *Ustilaginaceae* incluídos nas análises de genômica comparativa + *Malassezia globosa* (grupo externo).

Figura 15 - Prevalência de famílias de CAZymes em *Ustilaginaceae*.

Figura 16 - Clusterização hierárquica baseada no número de proteínas atribuídas às famílias de CAZymes (ou “perfil de CAZymes”) para os organismos usados nas análises comparativas.

Figura 17 - Clusterização hierárquica baseada no número de proteínas atribuídas às famílias de TAPs (ou “perfil de TAPs”) para os organismos usados nas análises comparativas.

Figura 18 - Cronograma de *Ustilaginaceae* com idades de nós em Milhões de Anos e evolução de famílias de CAZymes e TAPs. A árvore filogenética de *Ustilaginaceae* com genomas sequenciados, gerada com os quatro marcadores descritos por Capella-Gutierrez et al. (2014), foi submetida a uma análise usando o programa treePL para a predição de idade dos nós na árvore. Os nós internos usados para

calibração da árvore correspondem aos ancestrais comuns mais recentes a *M. maydis* e *A. flocculosa* (de 38 a 104 milhões de anos) e de *M. maydis* e *S. reilianum* (de 20 a 68 milhões de anos), estimado por Riess et al. (2016) e estão destacados em negrito (calibração secundária). O cronograma foi usado para a análise de expansão e contração de famílias com o programa CAFE v.4. São apresentadas as famílias que apresentaram alteração significativa na taxa de evolução nos nós internos (esquerda) e nos membros da família (direita). Em negrito, destacam-se as famílias expandidas e contraídas em *K. brasiliensis* e nas leveduras sequenciadas neste trabalho (vide Tabela 9 para consultar os nomes correspondentes aos identificadores).

Figura adicional 1 - Diagrama representando regras para anotação de Proteínas Associadas à Transcrição (TAPs) a partir de busca de domínios conservados (“in house” e do Pfam).

Figura adicional 2 - Filogenia de Máxima Verosimilhança (ML) com todas as espécies em *Ustilaginaceae* usando regiões ITS1, 5.8S, ITS2 e LSU com partições usando o programa RAxML, levando em conta a estrutura consenso gerada pelo mxscarna e o modelo S7A, modelo GTR para frequência de nucleotídeos e distribuição GAMMA, que leva em conta taxas variáveis entre as posições. Valores de suporte são apresentados para 1000 repetições de bootstrap. Edição final com enraizamento usando “mid-point root”.

LISTA DE TABELAS

Tabela 1 - Informações detalhadas dos desenhos experimentais de sequenciamento de transcriptoma com RNA-seq (“pool” e “fontes de carbono”).

Tabela 2 - Evidências de cada fonte que contribuiu para o melhoramento da anotação de *Kalmanozyma brasiliensis* GHG001.

Tabela 3 - Número de genes em cada grupo de perfis de expressão (apenas genes diferencialmente expressos).

Tabela 4 - Genes de *K. brasiliensis* GHG001 que codificam proteínas com domínios de CAZymes possivelmente envolvidas no metabolismo de xilano (Dodd e Cann, 2009; Peng et al., 2017) e que estão incluídos entre os diferencialmente expressos.

Tabela 5 - Identificação de transportadores putativos de xilose em *K. brasiliensis* GHG001.

Tabela 6 - Genes de *K. brasiliensis* GHG001 que codificam proteínas com ortólogos em reguladores caracterizados (Todd et al., 2014).

Tabela 7 - Genes identificados no secretoma de *K. brasiliensis* GHG001 em crescimento em diferentes fontes de carbono (xilose, xilano e glicose) e expressão no presente trabalho.

Tabela 8 - Análise de enriquecimento de mapas do KEGG e famílias de CAZymes e TAPs para os grupos de genes diferencialmente expressos.

Tabela 9 - Genomas de Ustilaginomycotina incluídos nas análises de filogenômica e genômica comparativa (basidiomicetos incluídos na figura 9 estão apresentados na tabela adicional 5).

Tabela 10 - Informações sobre os quatro marcadores de Capella-Gutierrez et al. (2014) e perfis HMM construídos para identificação nos proteomas incluídos neste trabalho.

Tabela 11 - Número de “reads” sequenciadas e limpas para o sequenciamento do genoma de leveduras isoladas de flores de *Heliconia psittacorum*: *Moesziomyces* spp. F16C1 e F5C1 e *Pseudozyma* sp. F8B2.

Tabela 12 - Estatísticas da montagem de genomas de leveduras isoladas de flores de *Heliconia psittacorum*: *Moesziomyces* spp. F16C1 e F5C1 e *Pseudozyma* sp. F8B2.

Tabela 13 - Análise com BUSCO para 1438 ortólogos de cópia única em fungos da família *Ustilaginaceae* incluídos neste trabalho.

Tabela adicional 1 - Proteínas de *K. brasiliensis* GHG001 anotadas com domínios de CAZymes possivelmente envolvidas no metabolismo de xilano.

Tabela adicional 2 - Estatísticas de mapeamento de *reads* no genoma de *K. brasiliensis* GHG001 para cada experimento.

Tabela adicional 3 - Número de CAZymes e TAPs preditas no proteoma melhorado de *K. brasiliensis* GHG001.

Tabela adicional 4 - Anotação de genes de *K. brasiliensis* associados a mapas do KEGG considerados enriquecidos nos diferentes grupos de perfil de expressão gênicas (DEG).

Tabela adicional 5 - Basidiomicetos (Agaricomycotina e Pucciniomycotina) incluídos na filogenia usando os quatro marcadores descritos em Capella-Gutierrez et al. (2014).

Tabela adicional 6 - Macrovariáveis associadas aos genomas usados nas análises comparativas (*Ustilaginaceae* + *M. globosa*).

Tabela adicional 7 - Resultados de expansão de contração de famílias de proteínas (CAZymes e TAPs) em *Ustilaginaceae*.

Tabela adicional 8 - Número de diferenças (substituições, inserções e deleções) em comparações das regiões ITS1, ITS2, 5.8S e subunidade maior ribossomal entre as leveduras sequenciadas neste trabalho e as mais próximas na análise filogenética (figura adicional 2).

Tabela adicional 9 - Número de componentes (genes, reações, metabólitos) nos modelos metabólicos usados neste trabalho.

LISTA DE ABREVIASÕES

AA - Auxiliary Activity

CAZyme - Carbohydrate-Active Enzyme (Enzima Ativa em Carboidrato)

CCR - Carbon Catabolite Repressor

CE - Carboidrato Esterase

CWDE - Enzimas de degradação de parede celular

FBA - Flux Balance Analysis (Análise de Balanço de Fluxo)

GH - Glicosil Hidrolase

GLUIG - Genes Induzidos em Glicose

GLURG - Genes Reprimidos em Glicose

ITS - Internal Transcribed Spacer

LPMO - Lytic Polysaccharide Monooxidase

MFS - Major Facilitator Superfamily (transportadores)

oTR - Outros Reguladores de Transcrição

PCM - Métodos Filogenéticos Comparativo

PL - Polissacarídeo Liase

PPP – Pentose Phosphate Pathway (Via de Pentoses Fosfato)

SP - Sugar Porter (transportadores de açúcar)

TAP - Proteína Associada à Transcrição

FT - Fator de Transcrição

XLAIG - Genes Induzidos em Xilano

XLARG - Genes Reprimidos em Xilano

XLOIG - Genes Induzidos em Xilose

XLORG - Genes Reprimidos em Xilose

XOS - Xilo-oligômero

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE ILUSTRAÇÕES	13
LISTA DE TABELAS	17
LISTA DE ABREVIATURAS	20
INTRODUÇÃO	23
REVISÃO BIBLIOGRÁFICA	27
<i>Kalmanozyma brasiliensis GHG001</i>	27
A família <i>Ustilaginaceae</i> e seu potencial biotecnológico	28
A degradação de xilano, o transporte e o metabolismo de xilose em fungos	29
Regulação da transcrição na degradação de xilano e metabolismo de xilose em fungos basidiomicetos	33
Taxonomia e filogenia de <i>Ustilaginaceae</i>	35
Genômica comparativa e análises incluindo <i>Ustilaginaceae</i>	37
OBJETIVOS GERAIS	39
OBJETIVOS ESPECÍFICOS	39
Capítulo 1 - Análises de transcriptoma de <i>Kalmanozyma brasiliensis GHG001</i> usando RNA-seq	40
INTRODUÇÃO.....	40
MATERIAIS E MÉTODOS	43
Cultivo de <i>K. brasiliensis GHG001</i> e sequenciamento do transcriptoma	43
Melhoramento da anotação estrutural de <i>K. brasiliensis GHG001</i>	44
Anotação funcional de <i>K. brasiliensis GHG001</i>	45
Análises de expressão diferencial de genes e enriquecimento	46
Disponibilização dos dados.....	47
RESULTADOS E DISCUSSÃO	47
O melhoramento da anotação do genoma de <i>K. brasiliensis GHG001</i>	47
A anotação funcional de <i>K. brasiliensis GHG001</i>	49
As análises de transcriptoma de <i>K. brasiliensis GHG001</i> em diferentes fontes de carbono	50
CONCLUSÕES.....	63
Capítulo 2 - Análises de genômica comparativa de <i>Ustilaginaceae</i> , incluindo o sequenciamento de novas leveduras da família	65
INTRODUÇÃO	65
MATERIAIS E MÉTODOS	69

Sequenciamento de novas leveduras da família <i>Ustilaginaceae</i>	69
Genomas usados na análise comparativa	70
Identificação de ortólogos nos proteomas	71
Reconstrução filogenética	71
Evolução de tamanhos de famílias gênicas.....	73
Disponibilização dos dados.....	73
RESULTADOS E DISCUSSÃO	73
O sequenciamento dos genomas de leveduras isoladas de flores de <i>H. psittacorum</i>	74
As filogenias de espécies e a alocação das novas leveduras sequenciadas	76
Distribuição de CAZymes e TAPs.....	83
Análises de expansão de contração de famílias gênicas em <i>Ustilaginaceae</i>	87
CONCLUSÕES.....	91
CONCLUSÕES E PERSPECTIVAS.....	93
REFERÊNCIAS BIBLIOGRÁFICAS.....	94
ANEXOS	109
Anexo 1 - Figuras adicionais	109
Anexo 2 - Tabelas adicionais.....	112
Anexo 3 - Lista de publicações relacionadas ao mestrado	144
Anexo 4 - Códigos usados no trabalho	148
Anexo 5 - Participação em eventos e apresentação de trabalhos	149
Anexo 6 - Contribuição da bolsa de estágio e pesquisa no exterior (BEPE-FAPESP)	149
Anexo 7 - Iniciativas educacionais	151
DECLARAÇÃO BIOÉTICA E BIOSSEGURANÇA.....	153
DECLARAÇÃO DE DIREITOS AUTORAIS	154

INTRODUÇÃO

Os fungos basidiomicetos (Filo Basidiomycota) são ecologicamente importantes, consistindo principalmente em sapróbios e fitopatógenos. O filo consiste em três subfilos: Pucciniomycotina (ferrugens), Ustilaginomycotina (fungos de carvão) e Agaricomycotina (cogumelos e parentes). O subfilo Ustilaginomycotina consiste de quatro classes: Exobasidiomycetes, Malasseziomycetes, Moniliellomycetes e Ustilaginomycetes. A classe Ustilaginomycetes é composta principalmente por fungos dimórficos que apresentam um estágio filamentoso patogênico e produzem teliósporos no tecido do hospedeiro, sendo *Mycosarcoma* (= *Ustilago*) *maydis*, patógeno de milho, o mais estudado (Spatafora et al., 2017). Dentro desta classe, a família *Ustilaginaceae* (Ustilaginales) é composta por espécies leveduriformes e fungos filamentosos que apresentam um estágio haploide leveduriforme, distribuídas em 14 gêneros, entre eles *Ustilago*, *Sporisorium*, *Kalmanozyma* e *Moesziomyces*, e cuja taxonomia e relações filogenéticas foram recentemente revisadas (Wang et al., 2015).

A importância da família é ampla. Além de *M. maydis*, outros patógenos de plantas pertencentes à família incluem *Sporisorium reilianum*, causador de carvão em milho e *Sporisorium scitamineum*, causador de carvão em cana-de-açúcar (*Saccharum* ssp.). *M. maydis*, em particular, tem importância relevante na ciência por ser um modelo em estudos em genética e biologia molecular (Kämper et al., 2006). Os membros da família *Ustilaginaceae* também têm sido explorados biotecnologicamente. As espécies fitopatogênicas têm recebido atenção na produção de metabólitos secundários, como glicolipídeos, e enzimas de degradação de biomassa lignocelulósica (Guevarra e Tabuchi, 1990; Feldbrügge et al., 2013; Geiser et al., 2014; Paulino et al., 2017), enquanto espécies não patogênicas como *Anthracocystis flocculosa* e *Moesziomyces aphidis* são agentes de controle biológico (Mimee et al., 2005; Gafni et al., 2015) e *Moesziomyces antarcticus* produz biossurfactantes (Faria et al., 2015).

Durante o “screening” de leveduras fermentadoras de xilose para a produção de bioetanol de segunda geração, o isolado GHG001 destacou-se como promissor na produção da endoxilanase *PbXynA* da família GH11 com alta atividade enzimática em polímeros de xilano (Borges et al., 2014). Após a caracterização de sua enzima, particularidades moleculares e fisiológicas permitiram descrevê-lo como uma nova espécie de levedura, a *Pseudozyma brasiliensis* GHG001 (de Castro

Oliveira et al., 2014). Recentemente, juntamente com a realocação de outras espécies, a levedura foi atribuída ao gênero *Kalmanozyma*, passando a ser identificada como *K. brasiliensis* GHG001 (Wang et al., 2015). Com o objetivo de explorar o potencial biotecnológico desta levedura, o seu genoma foi também sequenciado por nosso grupo de pesquisa (de Castro Oliveira et al., 2013). Desde então, outros projetos têm buscado explorar o potencial da levedura, em particular para a produção de bioetanol de segunda geração, como a análise do secretoma realizada por Kaupert Neto et al. (2015), que visou identificar as enzimas secretadas e que podem atuar na degradação da parede celular de plantas como a cana-de-açúcar.

A parede celular de plantas é constituída de três principais polímeros: a celulose, a hemicelulose e a lignina. Como o xilano é o principal componente da porção hemicelulósica, o melhor entendimento da desestruturação deste polímero é importante para melhorar a eficiência da conversão de biomassa lignocelulósica em biocombustíveis (Dodd e Cann, 2009). Após ter ocorrido a desestruturação do xilano, o aproveitamento do substrato resultante da digestão enzimática, a xilose, é possível caso os microrganismos sejam capazes de transportá-la para o interior celular (Sloothaak et al., 2016). Por fim, a assimilação desta xilose envolve processos de conversão em xilulose, seguidos de fosforilação, o que permite conectá-la com o metabolismo central pela Via de Pentoses Fosfato (PPP) (Hahn-Hägerdal et al., 2007).

Como *K. brasiliensis* é capaz de crescer em xilose e xilano (de Castro Oliveira et al., 2014) e que a ação da endoxilanase caracterizada gera xilo-oligômeros (Borges et al., 2014), é possível que esta levedura tenha outras enzimas envolvidas na degradação de xilanos, bem como transportadores eficientes para produtos de sua digestão, particularmente de xilose (outras enzimas, ainda não caracterizadas, podem ser responsáveis pela quebra subsequente de xilo-oligômeros). Em *K. brasiliensis* GHG001, além da endoxilanase *PbXynA*, foram caracterizados os genes codificantes de enzimas envolvidas no metabolismo interno de xilose, Xilose Redutase (XR, gene *xy1/1*) e Xilitol Desidrogenase (XDH, gene *xy1/2*) (Borges, 2014), provendo uma plataforma inicial para estudos do metabolismo de açúcares como o xilano.

A regulação e a coordenação de crescimento, desenvolvimento, progressão do ciclo celular, a adaptação fisiológica e metabólica às variáveis ambientais

dependem da atuação de proteínas que se ligam ao DNA, os Fatores de Transcrição (FTs). Os FTs controlam a ativação ou a repressão da expressão gênica. Adicionalmente, outras proteínas reguladoras da transcrição (oRTs) estão envolvidas através da intervenção em interações proteína-proteína ou no remodelamento da cromatina (Pérez-Rodríguez et al., 2009; Buitrago-Flórez et al., 2014). A ativação e a repressão da expressão gênica em genes envolvidos na degradação de biomassa também é regulada por estes elementos, já havendo muitos FTs caracterizados em fungos, principalmente em Ascomicetos (Amore et al., 2013; Benocci et al., 2017). Em basidiomicetos, a regulação de celulases e hemicelulases é ainda pouco estudada e aparentemente existe pouca conservação entre os diferentes organismos no reino (Benocci et al., 2017); ainda assim, a presença de homólogos em diferentes grupos de fungos não garante a manutenção da sua função devido a um fenômeno conhecido como “transcriptional rewiring” (Rokas e Hittinger, 2007).

Em estudos de genômica, a anotação funcional pode revelar o potencial de degradação da biomassa lignocelulósica de um organismo, bem como o repertório de possíveis reguladores e proteínas envolvidas no transporte dos produtos de hidrólise para o interior celular. Para a busca de domínios conservados em proteínas, os perfis de HMMs são muito utilizados e a implementação da suíte HMMER3 permite fazer buscas rapidamente em proteomas completos (Eddy, 2009). As Enzimas Ativas em Carboidratos, conhecidas como “CAZymes”, estão envolvidas na síntese, degradação e modificação de carboidratos. Elas são atribuídas a diferentes classes funcionais: as glicosil hidrolases (GHs), as esterases de carboidratos (CEs), as glicosil transferases (GHs), as liases de polissacarídeos (PLs), além de módulos de ligação a carboidratos (CBM), que incluem as famílias envolvidas na degradação de xilanos previamente descritas e para as quais HMMs estão disponíveis no dbCAN (Yin et al., 2012). Os HMMs podem também ser utilizados para buscar por domínios conservados de proteínas reguladoras da transcrição (TAPs) que, neste caso, envolvem regras com combinações de domínios que caracterizam as famílias de reguladores (Pérez-Rodríguez et al., 2009; Buitrago-Flórez et al., 2014). Além disso, recentemente HMMs foram também desenvolvidos para a busca por transportadores de açúcares, como de glicose (Sloothaak et al., 2015) e xilose (Sloothaak et al., 2016).

Nas análises de genomas individuais, como de *K. brasiliensis* GHG001, após ter realizado a anotação funcional, dados de expressão gênica em xilose e xilano

usando RNA-seq podem dar fortes evidências sobre a atuação dos produtos destes genes, os quais podem ser caracterizados com mais detalhes em laboratório. Da mesma forma, a anotação funcional de genomas de um grupo de interesse, como a família *Ustilaginaceae*, permitem o uso de métodos de genômica comparativa para tirar proveito destas informações para ter um melhor entendimento sobre a evolução de organismos deste grupo, bem como para o entendimento da evolução do metabolismo de polímeros (como de carboidratos) e das relações ecológicas (como a relação entre patógeno e hospedeiro durante a infecção) (Floudas et al., 2012; Gazis et al., 2016; Gan et al., 2016), até mesmo para exploração do potencial biotecnológico (Riley et al., 2016) e escolha de organismos a serem explorados em laboratório e, eventualmente, na indústria.

Neste contexto, este trabalho aplica abordagens “ômicas”, ferramentas valiosas para a exploração do potencial biotecnológico de membros sequenciados da família *Ustilaginaceae*. Para melhorar a anotação estrutural de *K. brasiliensis* GHG001 e estudar o metabolismo de xilano e de seu monômero, a xilose, por esta levedura, o **capítulo 1 (“Análises de transcriptoma de *Kalmanozyma brasiliensis* GHG001 usando RNA-seq”)** explora dados experimentais, em particular de transcriptômica usando RNA-seq. Tendo em vista o potencial biotecnológico de diversos membros da família (Guevarra e Tabuchi, 1990; Feldbrügge et al., 2013; Geiser et al., 2014; Paulino et al., 2017), no **capítulo 2 (“Análises de genômica comparativa de *Ustilaginaceae*, incluindo o sequenciamento de novas leveduras da família”)**, é explorada a genômica comparativa de membros da família com especial ênfase em CAZymes e TAPs. Esta etapa do trabalho incluiu o sequenciamento de novas leveduras dos gêneros *Moesziomyces* e *Pseudozyma* pro tempore (termo sugerido por Wang et al. (2015) para indicar que *Pseudozyma* será mantida temporariamente devido às incertezas sobre sua localização filogenética), bem como sua alocação na árvore filogenética usando métodos de filogenia molecular para explorar dados de proteoma completo. O trabalho permite gerar um melhor entendimento sobre o potencial biotecnológico destes organismos ainda pouco explorados.

REVISÃO BIBLIOGRÁFICA

***Kalmanozyma brasiliensis* GHG001**

A levedura basidiomicética anamórfica *Kalmanozyma brasiliensis* GHG001 pertence à família *Ustilaginaceae* (Ustilaginales, Ustilaginomycetes, Ustilaginomycotina) e foi inicialmente isolada do trato intestinal da larva de um inseto *Chrysomelidae* que parasitava a cana-de-açúcar durante um projeto que visava o “screening” de microrganismos que metabolizassem xilano ou seu monômero, a xilose. Ela se destacou pela capacidade de crescimento em xilose e xilano como únicas fontes de carbono (de Castro Oliveira et al., 2014). Além disso, a levedura também se destacou pela produção de uma endoxilanase da família GH11, *PbXynA*, com alta atividade específica comparada com a de outros fungos (Borges et al., 2014). Assim como a enzima caracterizada de *Pseudozyma hubeiensis* pro tempore (Adsul et al., 2009), a endoxilanase de *K. brasiliensis* GHG001 gera xilo-oligômeros (XOS) e apresenta baixa produção de monômeros de xilose.

A produção de endoxilanase por *K. brasiliensis* GHG001 destaca seu potencial na produção de XOS, produto que pode ter aplicação na indústria de produção de alimentos; além disso, XOS tem propriedades interessantes para a indústria farmacêutica, como atividade anti-inflamatória e antialérgica (Moure et al., 2006). Além disso, a produção de xilanases também tem destaque na produção de produtos de degradação de biomassa lignocelulósica que podem ser utilizados para a produção de biocombustíveis, como o bioetanol de segunda geração, através da hidrólise de polímeros como o xilano (Dodd e Cann, 2009). Por fim, foram caracterizados os genes envolvidos no metabolismo intracelular de xilose nesta levedura, a XR e a XDH (Thuanny A. Borges, 2014).

Tendo em vista o potencial biotecnológico de *K. brasiliensis* GHG001, o seu genoma foi sequenciado (de Castro Oliveira et al., 2013), possibilitando uma melhor exploração do seu potencial biotecnológico. O sequenciamento do genoma foi realizado usando a tecnologia Illumina e a sua montagem resultou em 45 scaffolds, totalizando 17,32 Mb de tamanho e conteúdo GC de 56,3%. A anotação estrutural do genoma foi realizada, resultando em 5.768 genes codificadores de proteínas, sendo que 2.361 (40%) apresentam produtos com alta similaridade quando comparados com proteínas no NCBI (nr). O melhoramento desta anotação inicial do genoma de *K. brasiliensis* usando dados experimentais, como de transcriptômica por

RNA-seq, possibilita corrigir modelos de genes, o que resulta em alteração do tamanho do proteoma e de sua qualidade, possibilitando a realização de análises globais (como de genômica comparativa) de forma mais precisa.

O sequenciamento do genoma de *K. brasiliensis* GHG001 possibilitou uma análise acurada do secretoma da levedura crescida em diversas fontes de carbono, como xilose e xilano (Kaupert Neto et al., 2015). O estudo mostrou que diversas enzimas de degradação de polímeros lignocelulósicos, as CAZymes, foram identificadas no secretoma. Neste trabalho, também foi verificado o melhor crescimento da levedura em xilano, xilose e glicose em relação à celobiose, reforçando a ideia que este microrganismo metaboliza os polímeros de pentose e seus monômeros. Tendo em vista este “background” já fornecido para *K. brasiliensis* GHG001, abordagens “ômicas” permitem melhorar ainda mais o entendimento sobre a biologia desta levedura, como o metabolismo de pentoses.

A família *Ustilaginaceae* e seu potencial biotecnológico

Ustilaginaceae é uma família de fungos basidiomicetos, cuja importância é atribuída principalmente a patógenos de plantas, os chamados “fungos de carvão”. Estes fungos geralmente infectam gramíneas e apresentam um ciclo celular que envolve células haploides, leveduriformes e com diferentes tipos de acasalamento (“mating-types”), cuja união gera hifas dicarióticas responsáveis pela infecção (Bakkeren et al., 2008). Alguns exemplos de patógenos estudados na família são *M. maydis*, responsável pela infecção de milho (*Zea mays*) (Gunther et al., 2008) e *Sporisorium scitamineum*, que infecta a cana-de-açúcar (Sundar, 2012), além de *S. reilianum*, que infecta o sorgo e o milho, e *Ustilago hordei*, um patógeno de cevada e aveia (Vánky, 2012).

Além dos fungos teleomórficos (com reprodução sexual), a família também engloba membros anamórficos, leveduriformes, que possuem brotamento e cuja a reprodução sexual é ausente ou desconhecida, além de serem incapazes de infectar plantas (Begerow et al., 2000; Kruse et al., 2016). Nesta dissertação, são consideradas “leveduras” tanto as formas haploides de fungos que apresentam o ciclo sexual como as dos anamórficos (Wang et al., 2015).

Além da importância como fitopatógenos, muitos fungos da família se destacam pelo potencial biotecnológico (Guevarra e Tabuchi, 1990; Feldbrügge et al., 2013; Geiser et al., 2014; Paulino et al., 2017), uma vez que alguns membros

produzem poliois, ácidos orgânicos e lípideos. Os poliois, como eritritol e manitol, são aplicados na indústria de adoçantes e na indústria farmacêutica. Entre ácidos orgânicos, estão incluídos ácido itacônico, usado na produção de resinas, plásticos e adesivos, ácido málico, usado na produção de alimentos (principalmente acidulantes), ácido succínico usado na indústria farmacêutica, como aditivos, solventes e polímeros. Os lípideos extracelulares também são destaque em fungos de carvão, como lípideos manosilerititol e ácido ustilágico, que apresentam propriedades biossurfactantes, podendo ser usados nas indústrias farmacêutica, de cosméticos e alimentos; alguns deles são também conhecidos pelas propriedades fungicidas (Geiser et al., 2014).

Entre as leveduras com potencial biotecnológico na família, estão incluídas *A. flocculosa*, que se destaca por ser agente de biocontrole (Avis e Bélanger, 2002; Mimee et al., 2005; Marchand et al., 2009; Teichmann et al., 2011), *P. tsukubaensis*, aplicada industrialmente na produção de itaconato (Moon et al., 2010), *Moesziomyces antarcticus*, produtora de glicolípideos biossurfactantes (Faria et al., 2015), *Langdonia jejuensis*, produtora de cutinases (Seo et al., 2007) e a produtora de lipase *P. hubeiensis* (Bussamara et al., 2012).

Além destes produtos de interesse, enzimas interessantes também já foram caracterizadas em diversos membros de *Ustilaginaceae*, entre elas algumas envolvidas na degradação de biomassa lignocelulósica. Exemplos incluem xilanases em *P. hubeiensis* NCIM 3574 (Adsul et al., 2009) e *M. antarcticus* T-34 e GB-4(0) (Watanabe et al., 2015), xilosidases por *P. hubeiensis* NCIM 3574 (Mhetras et al., 2016) e α-L-arabinofuranosidase (GH62) por *M. maydis* (Siguier et al., 2014).

Além de todas estas características, muitos membros da família incluem a forma leveduriforme, que possui muitas vantagens em processos industriais. Adicionalmente, destaca-se a capacidade de utilização de diversas fontes de carbono, incluindo monômeros e polímeros derivados de fontes de biomassa renovável, o que é importante para a biotecnologia sustentável (Geiser et al., 2014).

A degradação de xilano, o transporte e o metabolismo de xilose em fungos

Os principais componentes da parede celular de plantas incluem a celulose, a hemicelulose e a pectina; juntamente com a lignina, eles formam um complexo que promove a rigidez e a estrutura da planta, protegendo-a do ataque microbiano (Rytioja et al., 2014). As enzimas envolvidas na degradação destes polímeros são de

interesse biotecnológico, já que seus produtos podem ser precursores em diversos processos de geração de produtos de interesse, como na indústria de combustíveis, alimentos, papel e produtos químicos (Himmel et al., 2007). O xilano, em particular, é o componente da hemicelulose que constitui entre 20 e 40% da parede primária e entre 40 e 50% da parede secundária de monocotiledôneas e representa também entre 20 e 30% da parede secundária de dicotiledôneas (Scheller e Ulvskov, 2010). Como as monocotiledôneas, que incluem gramíneas, são consideradas as mais importantes culturas em energias renováveis (Vogel, 2008), o entendimento da degradação de xilano e o metabolismo dos produtos de degradação, os xilo-oligômeros e a xilose, é essencial para aproveitamento do potencial biotecnológico de fungos.

O xilano é constituído por monômeros de xilose na cadeia principal conectados por ligação β -1,4 e podendo apresentar grupos funcionais na cadeia lateral, como acetil, arabinofuranosil, além de possivelmente ter ligação com a lignina através de esteres aromáticos (Scheller e Ulvskov, 2010; Figura 1). Sua degradação envolve diferentes enzimas, incluindo endo-1,4- β -xilanases (E.C. 3.2.1.8), β -D-xilosidases (E.C. 3.2.1.37), α -L-arabinofuranosidases (E.C. 3.2.1.55), α -glucuronidases (E.C. 3.2.1.139), acetil xilano esterases (E.C. 3.1.1.72) e ácido coumárico/ ferúlico esterases (E.C. 3.1.1.73) (Dodd e Cann, 2009; Figura 2). Entre estas enzimas, as xilanases, por exemplo, têm diversas funções nas indústrias de papel, de alimentação animal, de vinhos e sucos, na produção de xilitol, entre outras (Polizeli et al., 2005), valendo destacar também a produção de xilo-oligômeros (Jayapal et al., 2013; Borges et al., 2014). Na família *Ustilaginaceae*, já foram identificadas enzimas da degradação de xilano, como as endoxilanases de *P. hubeiensis* e *M. maydis* (Adsul et al., 2009; Geiser et al., 2013), a β -xilosidase de *P. hubeiensis* (Mhetras et al., 2016), α -L-arabinofuranosidase de *M. maydis* (Siguier et al., 2014). Após ter sido realizada a quebra de xilano em xilose, esta pentose precisa ser internalizada para que possa ser utilizada pelos microrganismos.

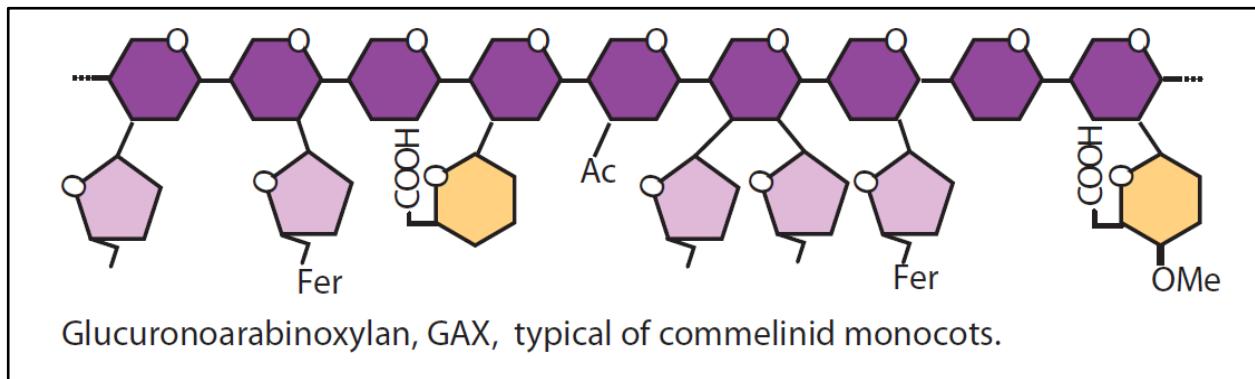


Figura 1 - Ilustração esquemática de um tipo de xilano que apresenta alta quantidade (% peso/peso, em relação a outros polímeros) na parede celular primária e secundária de gramíneas, o glucuronoarabinoxilano (imagem modificada de Scheller e Ulvskov, 2010).

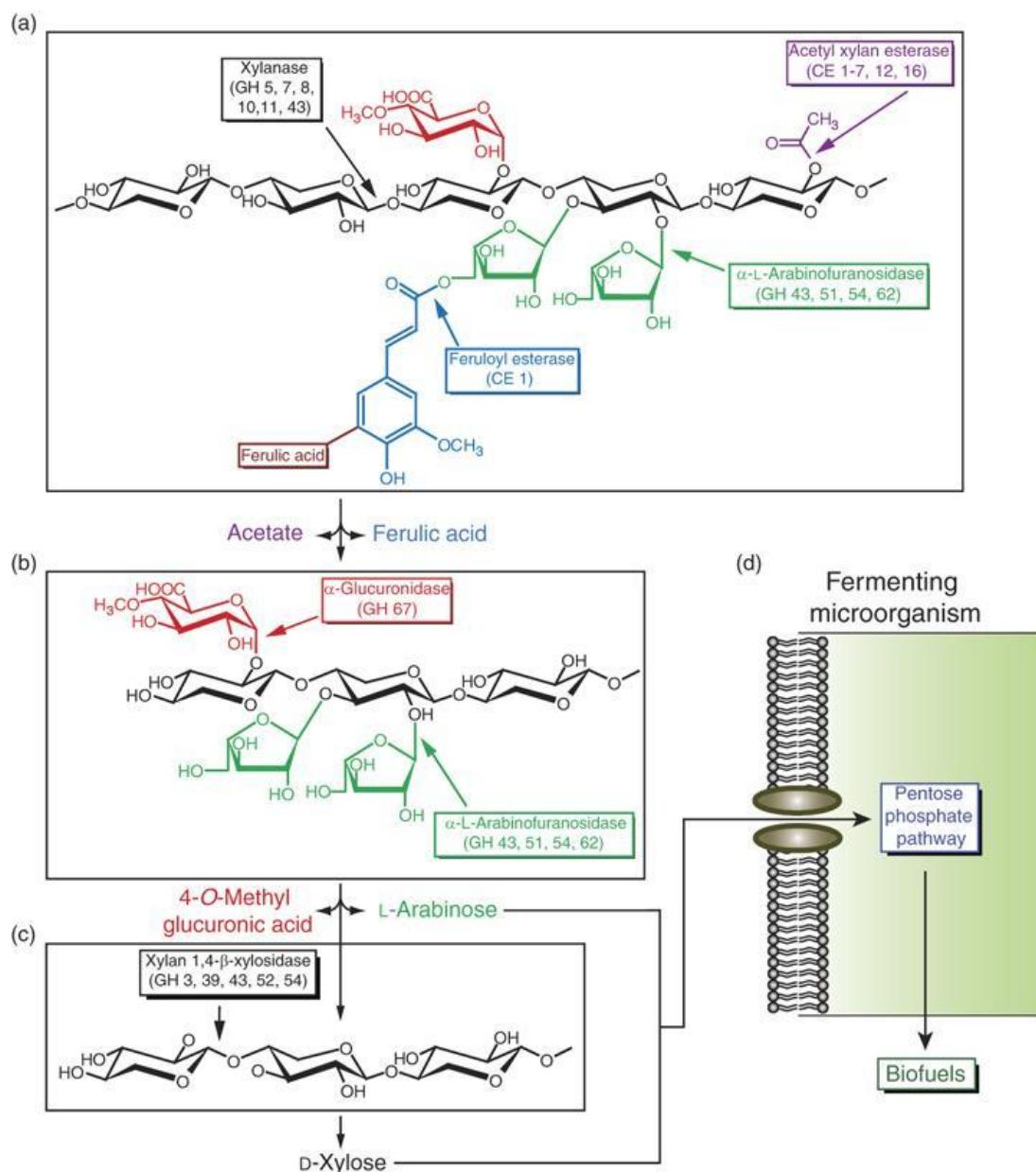


Figura 2 - Ação coordenada de enzimas para a desconstrução de xilano na produção de biocombustíveis. a) Xilanases, acetil xilano esterases e ácido ferúlico esterases atuam juntas para a

geração de pequenos xilo-oligossacarídeos, com a liberação de ácidos ferúlico e acético. b) Arabinofuranosidases e glucuronidases liberam arabinose e ácido glucurônico destes xilo-oligossacarídeos. c) Xilosidases convertem xilo-oligossacarídeos em xilose. d) Organismos fermentadores assimilam a xilose e a arabinose, direcionando as pentoses para a Via de Pentoses Fosfato para fermentação de biocombustíveis (imagem retirada de Dodd e Cann, 2009).

Uma das principais limitações do metabolismo de cepas industriais, como a levedura ascomicética *Saccharomyces cerevisiae*, para a produção de bioetanol usando pentoses é o transporte destes açúcares (como a xilose) para o meio intracelular. Em *S. cerevisiae*, sabe-se que existem inúmeros transportadores, muitos com função redundante e com forte preferência por glicose. Usando métodos computacionais de busca de domínios conservados do Pfam (PF07690), foram identificadas 73 proteínas pertencentes à “Major Facilitator Superfamily” (PF07690), sendo 43 pertencentes à “Sugar Porter” (PF00083) (Sloothaak et al., 2016). Em relação aos transportadores de xilose, a maior parte dos caracterizados são de ascomicetos, como o TrSTR1 em *Trichoderma reesei* (Huang et al., 2015), Gxs1 em *Candida intermedia* (Leandro et al., 2006), Sut1, Sut2 e Sut3 em *Scheffersomyces stipitis* (Weierstall et al., 1999) e XtrD em *Aspergillus nidulans* (Colabardini et al., 2014). Por outro lado, não há trabalhos caracterizando transportadores de xilose em basidiomicetos, limitando possíveis aplicações do potencial biotecnológico destes fungos e o entendimento do metabolismo de xilose nestes organismos.

Após a importação da xilose, é necessário a sua metabolização e eventual fermentação. O metabolismo de pentoses em fungos foi revisado por Hahn-Hägerdal et al. (2007) (Figura 3). Nestes organismos, a conversão de D-xilose para D-xilulose é geralmente realizada através de etapas de redução e oxidação envolvendo cofatores NAD(P)⁺/NAD(P)H e as enzimas xilose redutase XR e XDH (xilitol → D-xilulose). Por fim, a D-xilulose é fosforilada por uma xiluloquinase (XKS), gerando D-xilulose 5-fosfato, conectada ao metabolismo central pela Via de Pentoses Fosfato (PPP). A PPP, em particular sua porção não oxidativa, é responsável pela introdução de xilulose na via glicolítica através de reações envolvendo transaldolases e transacetolases. Estudos envolvendo o melhoramento de cepas industriais de *S. cerevisiae* para crescimento em xilose e consumo de xilulose visaram a superexpressão de enzimas desta porção não oxidativa da PPP (Matsushika et al., 2012).

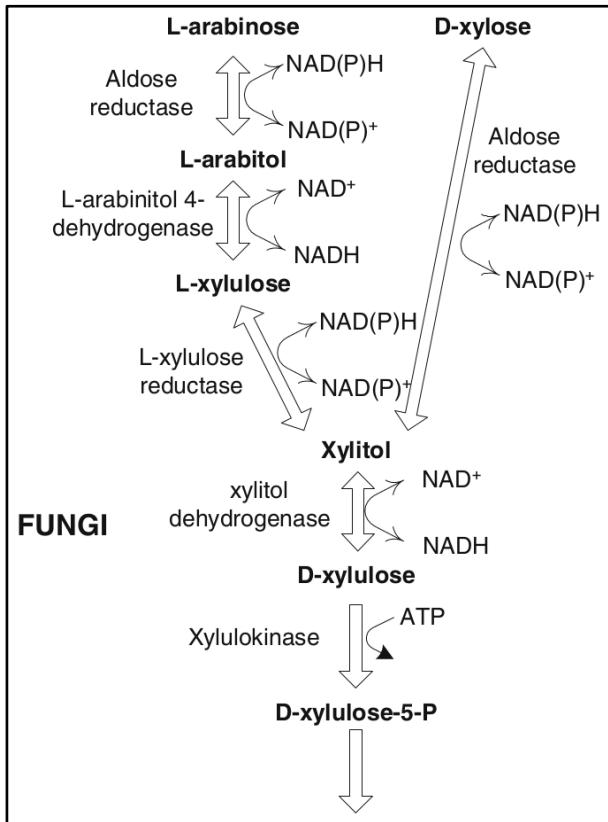


Figura 3 - Vias de utilização de xilose em fungos. Fonte: Modificado de Hahn-Hägerdal et al. (2007).

Neste contexto, abordagens “ômicas”, como a genômica, são promissoras para explorar o potencial de degradação de xilano através da busca por domínios de enzimas hidrolíticas ou de possíveis transportadores de xilose usando modelos computacionais. Além disso, métodos de transcriptômica usando RNA-seq são valiosos para identificar quais os genes estão sendo expressos ou reprimidos em condições de crescimento nestes açúcares e que estão possivelmente envolvidos na degradação de xilanos, metabolismo dos produtos (xilo-oligômeros e xilose) ou regulação da expressão de genes envolvidos na degradação e metabolismo interno.

Regulação da transcrição na degradação de xilano e metabolismo de xilose em fungos basidiomicetos

Os mecanismos de regulação de genes que codificam celulases e hemicelulases foram estudados principalmente nos fungos ascomicetos dos gêneros *Aspergillus* e *Trichoderma* e foram revisados recentemente (Amore et al., 2013; Benocci et al., 2017). Como a produção destas enzimas envolve gasto energético, existem mecanismos finos de regulação envolvendo indução ou repressão de seus respectivos genes. Em *T. reesei*, a indução de celulases e hemicelulases ocorre na

presença de polímeros (xilano, celulose e outros polímeros da parede celular de plantas) no meio de cultura, bem como por oligossacarídeos, como soforose, D-xilose, xilobiose, galactose e manose (Amore et al., 2013).

Algumas questões interessantes envolvendo regulação de CAZymes em organismos bem estudados incluem a inabilidade de indução da produção de enzimas por polímeros, levantando hipóteses de que enzimas de degradação de polímeros são expressas em níveis basais e são responsáveis pela liberação de indutores solúveis (Amore et al., 2013). Outras questões envolvem fatores transpcionais que co-regulam enzimas hidrolíticas e a possível atuação coordenada de alguns FTs, como reguladores positivos *XlnR*, *Ace2* e *HapB/C/E*, o repressor *Ace1* e *CreA* (repressor catabólico por carbono) (Amore et al., 2013).

A maior parte dos estudos supramencionados foi realizada em ascomicetos, sendo a regulação de enzimas envolvidas na degradação de biomassa pouco estudada em basidiomicetos. Além disso, Todd et al. (2014) buscaram 64 reguladores caracterizados de ascomicetos em diversos genomas de basidiomicetos e viram que muitos deles não apresentam ortólogos em basidiomicetos. Alguns já estudados, como o repressor catabólico por carbono *CreA*, no entanto, apresentam ortólogos em basidiomicetos.

De forma geral, poucos avanços têm sido observados no entendimento da regulação de enzimas da degradação de polímeros em basidiomicetos. Estudos com alguns deles relataram que há enzimas reprimidas por glicose e outros monossacarídeos, como foi observado em *Trametes versicolor* (“degradação branca”), *Agaricus bisporus* e *Volvariella volvacea*, o fitopatógeno *S. rolfsii* e a levedura *Rhodorotula minuta*. A repressão de uma endoxilanase por glicose também foi observada em *A. bisporus* (De Groot et al., 1998). Além da glicose, a inibição de genes que codificam enzimas hidrolíticas já foi observada na presença de monômeros de lactose, frutose, xilose e manose (Rytioja et al., 2014).

A indução de genes também já foi observada em poucos basidiomicetos. Enzimas hidrolíticas são induzidas em *F. palustris*, *G. trabeum* e *A. bisporus* quando expostos a polímeros de celulose e hemicelulose (Rytioja et al., 2014). Muitos mecanismos, no entanto, precisam ser ainda elucidados. Por exemplo, o motivo CCAAT está geralmente presente na região promotora de enzimas envolvidas na degradação de polímeros de celulose e hemicelulose em ascomicetos, mas está ausente em diversas enzimas de basidiomicetos, mostrando que os mecanismos

regulatórios podem ser diferentes e terem surgido paralelamente na evolução de ascomicetos e basidiomicetos (Rytioja, et al., 2014).

Um estudo recente mostrou o efeito de diversas fontes de carbono na expressão de enzimas hidrolíticas nos fungos basidiomicetos de degradação branca *Fomes fomentarius*, *Panus lecometei*, *Pseudotrametes gibbosa* e *Trametes versicolor* (Kobakhidze et al., 2016). Embora o estudo tenha mostrado que há respostas diferentes em termos de indução e repressão de enzimas nas diferentes fontes de carbono, estudos mais aprofundados são necessários para caracterização funcional dos mecanismos regulatórios.

Estudos de bioinformática envolvendo a busca por domínios conservados de proteínas TAPs, incluindo fatores de transcrição e outras proteínas envolvidas na regulação transcrional, como realizado por Pérez-Rodríguez et al. (2009) podem ser valiosos como ponto de partida para estudar a importância destes genes na regulação do metabolismo de carboidratos. Em especial, a identificação de TAPs através do emprego de regras com domínios de proteínas, sem a necessidade de busca por ortólogos, ameniza a dificuldade de trabalhar com um grupo no qual a regulação da expressão é ainda pouco estudada.

Taxonomia e filogenia de *Ustilaginaceae*

A taxonomia de fungos basidiomicetos anteriormente atribuía fungos sem o estado perfeito (aqueles cujas características morfológicas para a reprodução sexuada estão ausentes ou são desconhecidas) ao grupo de deuteromicetos, o que incluía as leveduras basidiomicéticas anamórficas da família *Ustilaginaceae*. Em 2000, Begerow incluiu estes fungos anamórficos em uma análise filogenética usando a subunidade maior da região ribossomal (rDNA) e mostrou que as leveduras anteriormente atribuídas ao gênero *Pseudozyma* formavam um grupo monofilético com fungos de carvão, como *Sporisorium* e *Ustilago*.

Apesar da monofilia caracterizar a família *Ustilaginaceae* dentro de Ustilaginales, as leveduras anteriormente consideradas “*Pseudozymas*” constituíam um grupo polifilético, distribuindo-se ao longo dos clados formados em filogenias da família, como observado por nosso grupo (de Castro Oliveira et al., 2014). Recentemente, Wang et al. (2015) realizou uma revisão de leveduras em Ustilaginales e atribuiu estas leveduras a diversos gêneros, entre eles: *Anomalomyces*, *Anthracobystis*, *Dirkmeia*, *Kalmanozyma*, *Langdonia*,

Macalpinomyces, *Melanopsichium*, *Moesziomyces*, *Sporisorium*, *Stollia*, *Tranzscheliella*, *Triodiomyces*, *Tubisorus*, e *Ustilago*. Nesta dissertação, a nomenclatura de membros de *Ustilaginaceae* segue a atribuição feita nesta publicação.

A região “Internal Transcribed Spacers” (ITS) do rDNA é considerada “barcode” em taxonomia de fungos (Schoch et al., 2012). No entanto, o uso de diversos outros marcadores já foi realizado para melhorar a acurácia da inferência. Riess et al. (2016) realizaram inferências filogenéticas usando supermatriz com dois grupos de genes: (i) 18S, 28S, rpb1 e (ii) 18S, ITS, 28S, rpb2, EF1 α . Na análise, foram incluídos diversos membros de Ustilaginomycotina, como *Moesziomyces bullatus*, *M. maydis*, *S. reilianum*, *A. flocculosa* como representativos da família *Ustilaginaceae*.

Além de análises usando marcadores amplificados, análises filogenômicas já foram realizadas incluindo membros da família *Ustilaginaceae*. Ortólogos de cópia única foram utilizados por Sharma et al. (2015) para reconstruir uma filogenia que incluiu *S. reilianum*, *M. maydis* e *U. hordei*. Nagy et al. (2015) reconstruíram uma filogenia de basidiomicetos, com foco particular em Agaricomycetes, e incluiu *M. maydis*. Dutheil et al. (2016) usaram ortólogos de cópia única para reconstruir uma filogenia incluindo *M. maydis*, *S. reilianum*, *U. hordei*, *S. scitamineum*, *A. flocculosa*, e *M. pennsylvanicum*. Recentemente, Ye et al. (2017) realizaram uma análise incluindo 45 ortólogos de cópia única em *Ustilago esculenta*, *M. maydis*, *U. hordei* e *S. reilianum*. Recentemente, Zhao et al. (2017) realizaram análise filogenética (filoproteômica) de basidiomicetos, incluindo todos os membros de *Ustilaginaceae* com genoma sequenciado, exceto *U. esculenta*. Capella-Gutierrez et al. (2014) utilizaram uma abordagem filogenômica para seleção de marcadores a serem usados na reconstrução de filogenias de espécies e quatro deles foram selecionados como ótimos para filogenias de espécies de fungos: “Target of rapamycin complex 1 subunit KOG1”, “Clustered mitochondria protein 1”, “Vacuolar protein sorting-associated protein 53” e “Replication factor A protein 1”. Desta forma, o uso desta abordagem ou a exploração de ortólogos de cópia única constituem formas de reconstruir uma filogenia acurada do grupo e alocar novos organismos sequenciados e permitir análises subsequentes de genômica comparativa.

Genômica comparativa e análises incluindo *Ustilaginaceae*

A genômica comparativa de fungos permite explorar diversos aspectos de sua biologia. Recentemente, um trabalho de genômica comparativa explorou fungos (ascomicetos do gênero *Aspergillus*) de importância médica e industrial, revelando diversidade funcional e adaptações específicas (de Vries et al., 2017). Em basidiomicetos, diversos estudos de genômica comparativa foram realizados para a análise de degradadores de biomassa, em particular o subfilo Agaricomycotina, pelo grande número de taxa em relação às outras classes de basidiomicetos e por serem tradicionalmente conhecidos como decompositores (Riley et al., 2014).

Nestes estudos, os fungos são classificados de acordo com o potencial de decaimento de biomassa, podendo ser de degradação branca, parda, branda ou “blue stain”. Os basidiomicetos, em particular, incluem fungos de degradação branca ou parda, sendo os primeiros especialmente eficientes na degradação de lignina (Riley et al., 2014; Lundell et al., 2014). É importante destacar que os organismos atribuídos a diferentes estratégias de degradação de biomassa não formam grupos monofiléticos (Ohm et al., 2014). As CAZymes abrangem as responsáveis pela síntese, modificação e degradação de carboidratos (Cantarel et al., 2008; Lombard et al., 2013) e análises de genômicas comparativa em fungos mostraram que há correlação entre o número e tipos destas enzimas e a habilidade de uso de diferentes fontes de carbono e substratos lignocelulósicos (Eastwood et al., 2011; Floudas et al., 2012).

Zhao et al. (2013) analisaram genômica comparativa de diferentes fungos com foco na degradação de biomassa lignocelulósica, incluindo a presença e ausência de famílias de CAZymes em ascomicetos e basidiomicetos. Entre as GHs, por exemplo, eles observaram que a distribuição e a abundância de CAZymes variam muito entre os diferentes fungos. Além disso, foi possível fazer um paralelo da presença de diversas famílias com o estilo de vida dos fungos, como a análise de pectinases e a relação com a patogenicidade dos organismos estudados.

Embora a maior parte dos estudos tenha sido realizada com ascomicetos ou basidiomicetos do subfilo Agaricomycotina, a disponibilidade de genomas sequenciados para fungos de carvão permitiu estabelecer relações de forma que não era possível antes. O patógeno mais estudado na família, *M. maydis*, teve seu genoma sequenciado em 2006 (Kämper et al., 2006), dando acesso, por exemplo, a diversos genes que poderiam estar relacionados com efetores durante a infecção,

sinalização de respostas a estresse, etc. Diversos outros patógenos foram sequenciados em seguida, incluindo *S. reilianum* (Schirawski et al., 2010), *Sporisorium scitamineum* (Que et al., 2014; Taniguti et al., 2015), *Ustilago hordei* Uh4857-4 (Laurie et al., 2012) e *U. esculenta* (Ye et al., 2017).

Entre os fungos anamórficos (leveduras) da família, diversos também já tiveram seus genomas sequenciados. O sequenciamento do genoma de *A. flocculosa* (Lefebvre et al., 2013), por exemplo, contribuiu para o entendimento da evolução de fungos de carvão, através do estudo da transição de uma forma patogênica para uma não infecciosa (*A. flocculosa*). Outros fungos foram sequenciados com uma justificativa biotecnológica: as leveduras *Moesziomyces antarcticus* (Morita et al., 2013; Saika et al., 2014), *Moesziomyces aphidis* (Lorenz et al., 2014) e *Pseudozyma hubeiensis* (Konishi et al., 2013), por exemplo, destacam-se pela produção de lipídeos. Entre elas, a levedura sequenciada por nosso grupo de pesquisa, *K. brasiliensis*, sequenciada tendo em vista a produção da enzima endoxilanase (de Castro Oliveira et al., 2013).

Entre os trabalhos de genômica comparativa de basidiomicetos, está incluída a evolução da degradação da biomassa por Floudas et al. (2015), que focou em membros de Agaricales (Agaricomycotina). Devido à importância como fitopatógenos, a genômica comparativa de Pucciniomycotina também foi explorada, com foco especial no estilo de vida de biotrófico obrigatório de membros como *Puccinia graminis* (Duplessis et al., 2011). No entanto, nenhum trabalho até o momento explorou a genômica comparativa incluindo um grande número de membros sequenciados de Ustilaginomycotina. Recentemente, Berlemont (2017) explorou enzimas de degradação de polissacarídeos nos fungos disponíveis no MycoCosm, no qual é evidenciado o número baixo de genomas deste grupo de fungos. A família *Ustilaginaceae* apresenta membros representativos e que dão oportunidade de explorar diferentes aspectos do grupo.

OBJETIVOS GERAIS

1. Melhorar a anotação estrutural do genoma de *Kalmanozyma brasiliensis* GHG001.
2. Analisar o transcriptoma de *K. brasiliensis* GHG001 em xilano e em xilose.
3. Estudar a genômica comparativa de *Ustilaginaceae* para contextualização do potencial biotecnológico de *K. brasiliensis* GHG001 e outras leveduras da família.

OBJETIVOS ESPECÍFICOS

1. Incorporar dados experimentais, principalmente de RNA-seq, para o melhoramento da anotação estrutural do genoma de *K. brasiliensis* GHG001.
2. Realizar a anotação funcional do proteoma melhorado de *K. brasiliensis* GHG001, principalmente de CAZymes, proteínas reguladoras da transcrição e transportadores de xilose.
3. Estudar o transcriptoma do crescimento de *K. brasiliensis* GHG001 com foco na anotação funcional realizada, em especial de enzimas relacionadas com o metabolismo de xilano e metabolismo de xilose, bem como o enriquecimento de mapas do KEGG, CAZymes e TAPs, para gerar *insights* sobre o metabolismo de pentoses na levedura.
4. Montar o genoma de leveduras basidiomicéticas isoladas de flores de *Heliconia psittacorum*.
5. Realizar inferências filogenéticas de espécies para alocação dos organismos sequenciados na filogenia de *Ustilaginaceae* e para as análises de genômica comparativa.
6. Realizar a anotação de CAZymes e TAPs em *Ustilaginaceae* com genomas sequenciados.
7. Estudar contração e expansão de famílias de CAZymes e proteínas reguladoras da transcrição em *Ustilaginaceae*, em especial de *K. brasiliensis* e das novas leveduras sequenciadas.

Capítulo 1 - Análises de transcriptoma de *Kalmanozyma brasiliensis* GHG001 usando RNA-seq

INTRODUÇÃO

As análises de transcriptômica possibilitam uma compreensão de diversos aspectos da biologia e responder a perguntas sobre o metabolismo em diferentes fontes de carbono. As tecnologias de alta performance, como o sequenciamento de transcritos usando RNA-seq, são de uso cada vez mais comum e não apresentam diversas limitações de tecnologias anteriores. Em relação aos microarranjos, por exemplo, o RNA-seq apresenta maior sensibilidade e menor variação técnica. Além disso, os resultados de RNA-seq apresentam uma forte correlação com resultados de microarranjos, tanto em valores absolutos como em análises de expressão diferencial (Oshlack et al., 2010), mostrando que a substituição do último é possível.

Diversas análises são possíveis usando dados de RNA-seq, entre elas a detecção de processamento alternativo e edição de RNAs, bem como a descoberta de novos transcritos. No entanto, a maioria dos trabalhos realizados empregando a tecnologia envolve perguntas sobre a diferença de expressão entre diferentes amostras. Alguns exemplos incluem a diferença de expressão entre linhagens mutantes e selvagens de um mesmo tecido, células tratadas e não tratadas, células tumorais *versus* tecido sadio, entre outras possibilidades (Oshlack et al., 2010). Para estudar o metabolismo de açúcares em diferentes condições, como na presença de xilano e xilose, o foco de estudo desta dissertação, é interessante também comparar a expressão com uma fonte de carbono que geralmente não induz a produção de enzimas, como a glicose.

Os fungos basidiomicetos, principalmente os membros do subfilo Agaricomycotina, são os principais decompositores de biomassa lignocelulósica (celulose, hemicelulose e lignina) (Lundell et al., 2014). As análises usando abordagens ômicas, como transcriptômica com RNA-seq, já foram feitas para entender o metabolismo de substrato lignocelulósico por diversos basidiomicetos, como *Pycnoporus sanguineus* (Rohr et al., 2013) e *Phlebia radiata* (Kuuskeri et al., 2016). Um exemplo de fungo cuja a maquinaria para decomposição de xilanos já foi estudada é o *Agaricus bisporus* (Agaricales) (Morin et al., 2012; Patyshakuliyeva et al., 2013).

Na família *Ustilaginaceae* (subfilo Ustilaginomycotina), o organismo mais explorado é o patógeno biotrófico facultativo de milho *M. maydis*, para o qual a maior parte dos estudos envolve a degradação de parede celular da planta durante o processo de infecção (Goulet e Saville, 2017). Análises de RNA-seq também já foram realizadas para estudar a interação de *Sporisorium scitamineum* e cana-de-açúcar(Que et al., 2014), ou para estudar a produção de lipídeos pela levedura da família *Moesziomyces aphidis* (Günther et al., 2015). Apesar de *K. brasiliensis* GHG001 ter se destacado pela produção da endoxilanase *PbXynA* (Borges et al., 2014) e na secreção desta e de outras CAZymes durante o crescimento em diferentes fontes de carbono, incluindo xilano e xilose (Kaupert Neto et al., 2015), nenhum estudo até o momento envolveu a análise de expressão diferencial em xilano e da xilose, visando o entendimento do metabolismo destes açúcares nesta levedura.

Os dados de RNA-seq, dependendo o desenho e planejamento experimental, têm potencial para responder a perguntas como: quais as enzimas de degradação de xilano induzidas neste polímero? Quais delas são reprimidas? Quais destas enzimas são induzidas em monômeros, como a pentose que constitui o xilano ou mesmo em glicose, possibilitando uma compreensão sobre a possível regulação por carbono? Quais fatores de transcrição previamente caracterizados em fungos estão induzidos ou reprimidos nas diferentes condições? Os resultados das comparações podem responder a algumas destas perguntas e constituir uma importante plataforma funcional para futuras validações em laboratório, conhecimento que pode ser aplicado em projetos de biotecnologia.

Para que as conclusões sobre os dados experimentais e resultados de análises *in silico* se aproximem mais da realidade biológica, projetos de anotação de genoma precisam investir na definição do ‘transcriptoma funcional’, a proporção do transcriptoma que contribui para a função celular. A anotação de genes envolve a determinação de modelos de transcritos, e a determinação da estrutura destes transcritos com a sua potencial função, entre as possíveis, se o mesmo corresponde a um gene codificador (Mudge e Harrow, 2016). Embora as possibilidades de funções dos transcritos sejam amplas, como os de pseudogenes, RNAs longos, famílias de RNAs pequenos como RNAs transportadores, PIWI-interacting RNAs, entre outros, o investimento em dados experimentais, recursos humanos e tempo

para o melhoramento da anotação de genomas depende muito da pergunta biológica em questão (Mudge e Harrow, 2016).

A Figura 4 apresenta estratégias de anotação para diferentes tipos de genes. No caso de genes em genomas não considerados “referência”, a melhor estratégia para a anotação de genes codificadores de proteínas envolve a incorporação de dados de homologia de proteínas em organismos relacionados e dados de RNA-seq (Figura 4 - b).

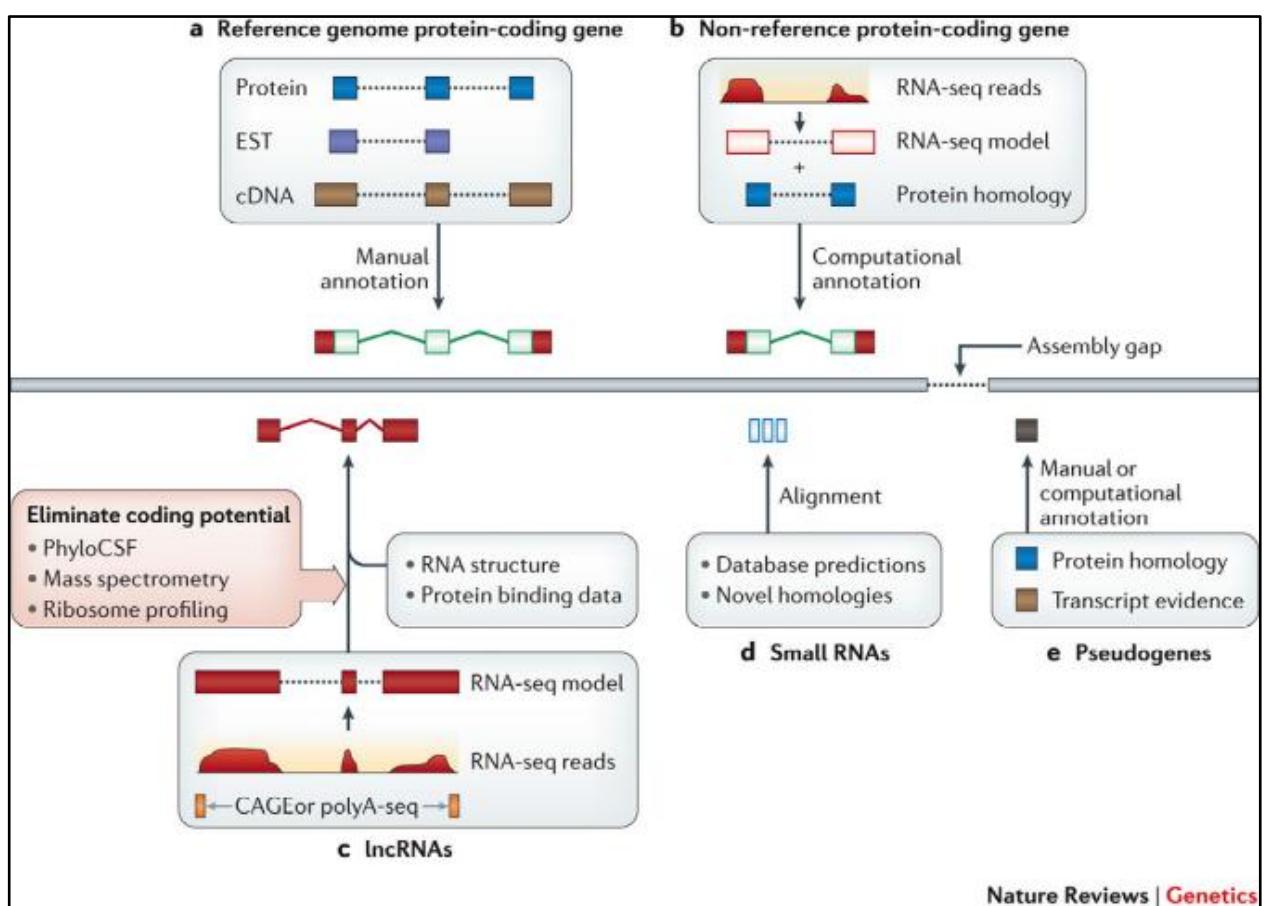


Figura 4 – Fluxo de trabalho de anotação para diferentes tipos de genes. a - Genes codificadores de proteínas em genomas de referência. b - Genes codificadores de proteínas em genomas não considerados referência. c - RNAs não codificadores longos. d - Pequenos RNAs. e - Pseudogenes. Fonte: Mudge e Harrow (2016).

Neste capítulo, são descritas análises de transcriptômica usando RNA-seq para o melhoramento da anotação (porção codificante de proteínas) de *K. brasiliensis* GHG001 e o entendimento do metabolismo de açúcares, em particular de xilano, um dos principais componentes da biomassa lignocelulósica, e de seu monômero, a xilose, através da análise de expressão diferencial comparando com crescimento em glicose.

MATERIAIS E MÉTODOS

Cultivo de *K. brasiliensis* GHG001 e sequenciamento do transcriptoma

Os experimentos de RNA-seq foram divididos em dois desenhos experimentais: “pool” e “fontes de carbono”. O pré-crescimento de *K. brasiliensis* GHG001 foi realizado em meio YPD (2% peptona, 1% extrato de leveduras, 4% dextrose, 1% ágar) a 30°C, por 24 horas, 10⁶ células/mL. Para os experimentos do “pool”, o inóculo foi realizado nas diferentes condições, variando temperatura, tempos de crescimento, meios de cultivo e fontes de carbono (incluindo “starvation”), de acordo com a Tabela 1, com o objetivo de maximizar a população de transcritos da levedura para melhoramento de sua anotação estrutural. Para os experimentos em “fontes de carbono”, a transferência foi feita para meio mínimo (YNB 0,67%) com diferentes fontes de carbono (1%): glicose, xilose ou xilano (BEECHWOOD da Sigma), por 24 horas, a 200 rpm, 30°C. Os experimentos foram realizados em triplicata. O preparo de bibliotecas foi realizado usando TruSeq, seguido de sequenciamento na plataforma Illumina HiSeq2000 do USC Epigenome Center (Los Angeles, Califórnia, EUA). Foram obtidas “reads paired-end” (pareadas) e “unstranded” (sem informação sobre a fita de DNA de origem) para ambos os experimentos.

Tabela 1 - Informações detalhadas dos desenhos experimentais de sequenciamento de transcriptoma com RNA-seq (“pool” e “fontes de carbono”).

Desenho experimental	Meio	Tempo de crescimento (horas)	Fonte de carbono	Temperatura (°C)
“Pool”	YNB	72	Arabinose	30
“Pool”	YNB	72	Cellobiose	30
“Pool”	YNB	72	Maltose	30
“Pool”	YNB	72	Sacarose	30
“Pool”	YNB	72	Galactose	30
“Pool”	YPD	96	Glicose	30
“Pool”	YPD	72	Glicose	30
“Pool”	YPD	48	Glicose	30
“Pool”	YPD	24	Glicose	30
“Pool”	YNB	48	Arabinose	30

“Pool”	YNB	48	Celobiose	30
“Pool”	YNB	48	Maltose	30
“Pool”	YNB	48	Galactose	30
“Pool”	YNB	48	Sacarose	30
“Pool”	YNB + casaaminoacidos	96	Glicose	30
“Pool”	YNB + casaaminoacidos	72	Glicose	30
“Pool”	YNB + casaaminoacidos (2)	48	Glicose	30
“Pool”	YNB + casaaminoacidos	48	Glicose	30
“Pool”	YNB + casaaminoacidos	24	Glicose	30
“Pool”	Pb SC	12	Glicose	30
“Pool”	Pb SC	24	Glicose	30
“Pool”	Pb MMS 0.005%	0,5	Glicose	30
“Pool”	Pb H2O2 20mM	0.25	Glicose	30
“Pool”	Pb YNB 2%	48	Glicose	30
“Pool”	YNB	0,5	-	37
“Pool”	YNB	1	-	37
“Pool”	YNB	0,5	-	42
“Pool”	YNB	1	-	42
“Fontes de Carbono”*	YNB	24	Glicose (GLU)	30
“Fontes de Carbono”*	YNB	24	Xilose (XLO)	30
“Fontes de Carbono”*	YNB	24	Xilano (XLA)	30

* Experimentos com fontes de carbono bem definidas foram realizados em triplicatas.

Melhoramento da anotação estrutural de *K. brasiliensis* GHG001

A anotação estrutural de genes foi previamente realizada em ocasião da montagem do genoma de *K. brasiliensis* GHG001 (de Castro Oliveira et al., 2013; montagem do NCBI GCA_000497045.1). O melhoramento da anotação estrutural foi realizado com a incorporação de dados de RNA-seq (ambos os desenhos experimentais). A qualidade do sequenciamento foi verificada com o programa FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Os programas BLAST (Altschul et al., 1990) e MEGAN5 (Huson e Mitra 2012) foram usados para verificar possíveis contaminações nos dados. A limpeza de “reads”, tal como a remoção das sequências curtas, com baixa qualidade e de regiões correspondentes

a sequências técnicas (como “primers”), foi realizada no Trimmomatic (Bolger et al., 2014). O mapeamento de “reads” foi realizado com Tophat 2.1 (Kim et al., 2013). O programa Qualimap 2 (Okonechnikov, Conesa et al. 2016) foi usado para verificar mapeamento de “reads” em regiões intrônicas. Por fim, a anotação estrutural incorporando informações de mapeamento de RNA-seq foi gerada com o Braker1 (Hoff et al., 2015). Dados de espectrometria de massas obtidos em outros projetos envolvendo *K. brasiliensis* (Borges et al., 2014; Kaupert Neto et al., 2015) e informações de homologia com outros *Ustilaginaceae* disponíveis na plataforma Mycocosm do JGI (Grigoriev et al., 2013) também foram incorporados na anotação final de genes codificadores de proteínas no programa EVidence Modeler (Haas et al., 2008). Os dados de espectrometria foram processados de acordo com a descrição a seguir: os diretórios RAW correspondentes aos espectros de massa brutos do Ultima Q-TOF (Waters) gerados por Kaupert Neto et al. (2015) foram convertidos para o formato MGF usando MS CONVERT (French et al. 2014). As sequências de peptídeos foram geradas *de novo* com o PepNovo+ (Frank e Pevzner, 2005) e a busca no genoma foi realizada usando o BLAST (tblastn) (Altschul et al., 1990). O programa Exonerate (Slater e Birney, 2005) foi usado para fazer a busca por homologia, incluindo a predição de íntrons, usando os proteomas de *Ustilaginaceae*. O programa BLAT v. 36x2 (Kent, 2002) foi usado para comparar as versões antiga e nova do proteoma de *K. brasiliensis* GHG001.

Anotação funcional de *K. brasiliensis* GHG001

Uma anotação geral do proteoma atualizado de *K. brasiliensis* GHG001 foi obtida usando o programa InterProScan 5 (Jones et al., 2015). Transportadores putativos de xilose foram obtidos com HMMER 3 (Eddy, 2009), usando HMMs desenvolvidos por Sloothaak et al. (2016) e aplicando como limiar de pontuação o menor valor descrito no mesmo artigo (472,9) para *A. niger* e *T. reesei*. Domínios de CAZymes foram identificados usando perfis HMM disponíveis no banco de dados dbCAN (Yin et al., 2012) e Proteínas Associadas à Transcrição (TAPs) foram anotadas usando regras descritas em Pérez-Rodríguez et al. (2009). Para as TAPs, regras para anotação haviam sido criadas para plantas e oomicetos (Pérez-Rodríguez et al., 2009; Buitrago-Flórez et al., 2014) e regras adicionais foram incluídas durante o desenvolvimento deste trabalho para FTs em fungos (Shelest 2008; de Mendoza et al., 2013; Todd et al., 2014) (figura adicional 1).

KAAS (Moriya et al., 2007), Blast2GO (Conesa et al., 2005) e PRIAM (Profls pour l'identification automatisée du métabolisme; Claudel- Renard et al., 2003) foram usados para a anotação de genes usados nas análises de enriquecimento, já que estes métodos permitem atribuição a mapas do KEGG. O rascunho do modelo de redes metabólicas de *K. brasiliensis* no formato SBML3 foi gerado usando PRIAM como ponto de partida para melhoramento e uma curadoria manual foi realizada durante o estágio realizado na Universidade de Virginia (Bolsa de Estágio e Pesquisa no Exterior; FAPESP nº 2016/13968-1). Parte da anotação funcional incluída no modelo metabólico foi transferida da anotação incluída no modelo metabólico de *M. maydis* 521, gentilmente fornecido pelo M.Sc. Christian Lieven e pelo Dr. Thiemo Zambanini, Alemanha. O rascunho do modelo gerado é apresentado no anexo 6.

Análises de expressão diferencial de genes e enriquecimento

As análises de expressão diferencial foram realizadas usando a versão atualizada da anotação estrutural de *K. brasiliensis* GHG001 (após incorporação de evidências experimentais com RNA-seq). “Reads” limpas foram mapeadas no genoma usando Tophat 2.1 (Kim et al., 2013). A quantificação da expressão foi realizada através da contagem de reads mapeadas em cada gene usando a função FeatureCounts implementada no pacote de R Rsubread (Liao et al., 2013). A matriz de contagem de “reads” por gene em cada condição foi usada como entrada nas análises subsequentes. Foram usados diferentes métodos para a análise de expressão diferencial: Counts + TMM (Robinson and Oshlack, 2010), Counts + Voom (Law et al., 2014) e Counts + Voom + Quantile (Hansen et al., 2012), como implementados nos DESeq2 (Love et al., 2014) e edgeR (Robinson et al., 2010). As análises subsequentes (lista de genes diferencialmente expressos, “clustering”, análises de enriquecimento) foram realizadas com a intersecção dos resultados dos diferentes métodos.

Os genes considerados diferencialmente expressos em pelo menos um contraste foram submetidos a uma análise de “clustering”. Para estes genes, os valores normalizados em FPKM foram convertidos para Z-scores e clusterizados usando o pacote de R Mclust (Scrucca et al., 2016). Os genes foram agrupados de acordo com os perfis de expressão: induzidos ou reprimidos em cada uma das fontes de carbono estudadas (glicose, xilose ou xilano).

Análise de enriquecimento de mapas do KEGG foi realizada para cada contraste e para cada grupo de perfis de expressão. Para tanto, mapas do KEGG associados (excluindo vias relacionadas a doenças humanas) a cada gene foram recuperados com um script “in house” Perl e KEGG API (Kawashima et al., 2003) a partir de identificadores do KEGG Orthology (KOs) e números EC, que por sua vez foram obtidos usando KAAS (Moriya et al., 2007), Blast2GO (Conesa et al., 2005) e PRIAM (Profils pour l'identification automatisée du métabolisme; Claudel- Renard et al., 2003). Uma função em R (anexo 4) foi usada para calcular o p valor associado a cada mapa e os resultados com p valor =< 0,05 foram considerados enriquecimentos significativos. A mesma função foi utilizada para verificar o enriquecimento de famílias de CAZymes e TAPs.

Disponibilização dos dados

Os dados brutos de sequenciamento de RNA-seq estão disponíveis no banco de dados SRA do NCBI associado ao BioProject PRJNA401334.

RESULTADOS E DISCUSSÃO

O melhoramento da anotação do genoma de *K. brasiliensis* GHG001

Neste trabalho, dois desenhos experimentais foram planejados para as análises de transcriptoma: (i) misturar diferentes tipos de induções para a produção de RNA (denominado “pool”) e (ii) diferentes “fontes de carbono” bem definidas (Tabela 1). Enquanto o último foi realizado em condições experimentais padronizadas (meio de cultivo, tempo de crescimento e temperatura), variando apenas as fontes utilizadas (glicose, xilose e xilano), o desenho do “pool” incluiu experimentos com diversos meios de cultivo, tempos de crescimento, temperaturas, fontes de carbono e em privação de carbono (“starvation”).

Para cada experimento de RNA-seq, as “reads” foram mapeadas no genoma de *K. brasiliensis* GHG001, previamente montado por nosso grupo de pesquisa e disponível no NCBI sob o identificador “GCA_000497045.1” (Tabela adicional 2). Além do RNA-seq, foram também incluídos os dados de espectrometria de massas obtidos anteriormente para a caracterização de uma endoxilanase da família GH11 (*PbXynA*; Borges et al., 2014) e no estudo do secretoma de *K. brasiliensis* GHG001 crescida em diferentes fontes de carbono (Kaupert Neto et al., 2015). Por fim, foi

realizada uma busca por homologia usando proteomas de *Ustilaginaceae* disponíveis do JGI. Todas estas evidências, apresentadas na tabela 2, foram incorporadas no programa EVidence Modeler, que gerou uma anotação final.

Tabela 2 - Evidências de cada fonte que contribuiu para o melhoramento da anotação de *K. brasiliensis* GHG001.

Método	Tipo de evidência	Número de evidências	Genes na anotação final
Anotação no NCBI (2013)	-	5.765 ***	5.765
Busca por homologia (Proteomas no JGI e Exonerate)	Hit de similaridade	57.571 ***	5.700
Sequenciamento de peptídeos (dados de espectrometria de massas) do secretoma *	Peptídeo	325 ***	81
Sequenciamento de peptídeos (spectromedia de massas) da endoxilanase PbXynA **	Peptídeo	10 ***	1
Experimentos de RNA-seq	Mapeamento de “read”	****	6.419
Anotação final (EVidence Modeler)	-	-	6.421

* Kaupert Neto et al., 2015

** Borges et al., 2014

*** IntersectBed foi usado para obtenção das intersecções entre arquivos de anotação (GFF) gerado para cada tipo de dado em relação à anotação final (EVidence Modeler)

**** Genes com o mínimo de dez “reads” mapeadas foram consideradas com evidência de RNA-seq

O proteoma final de *K. brasiliensis* GHG001 apresenta 6.421 proteínas, um proteoma com 656 proteínas adicionais quando comparado com 5.765 presentes no proteoma gerado em ocasião da montagem do genoma. No entanto, uma busca usando BLAT revela 817 proteínas encontradas exclusivamente na nova anotação, ou seja, além de descoberta de novos genes, a estrutura daqueles previamente existentes também mudou, havendo correções como na mudança da fase de leitura e na disposição de exons.

Como o desenho experimental de “pool” foi planejado exclusivamente para o melhoramento da anotação estrutural, foi analisada a contribuição destes dados na

anotação final. O mapeamento de “reads” de RNA-seq no genoma permite a geração de uma matriz de contagem por gene em cada uma das condições experimentais (glicose, xilose, xilano e “pool”) e a análise desta matriz permite a obtenção deste objetivo de anotação. Foram selecionados os genes que apresentaram média de contagem de “reads” abaixo de 50 para cada um dos experimentos em fontes de carbono e acima deste limiar no “pool”. Foram identificados 124 genes com contribuição do “pool”, justificando o uso destes dados como evidências para melhoramento da anotação.

A anotação funcional de *K. brasiliensis* GHG001

A anotação funcional do proteoma melhorado de *K. brasiliensis* revelou 5.934 proteínas com domínios conservados, identificados usando o pipeline do InterProScan 5 (ou seja, 487 proteínas não apresentam domínios conservados). Entre as proteínas com domínios conservados, 5.139 estão associadas a uma assinatura do InterPro (IPR). Entre as proteínas sem IPR, 486 estão associadas predominantemente a identificadores do Phobius, nas quais são identificadas regiões associadas à membrana ou a peptídeo sinal; 291 proteínas estão associadas a domínios do TMHMM (proteínas associadas à membrana); 79 proteínas estão associadas a um domínio conservado no Pfam.

Adicionalmente à anotação com InterProScan, foi realizada a anotação funcional de CAZymes e TAPs. *K. brasiliensis* GHG001 apresenta 262 proteínas atribuídas a 99 famílias de CAZymes e 296 proteínas atribuídas a TAPs. Foram também anotados genes de interesse no transporte e metabolismo de xilose. Sete transportadores putativos de xilose foram identificados usando perfis HMMs desenvolvidos por Sloothaak et al. (2016) e as enzimas Xilose Redutase (*xy1*) e Xilitol Desidrogenase (*xy12*), previamente caracterizadas por Borges (2014) foram identificadas no proteoma melhorado usando BLAST e correspondem aos genes evm.model.KI545895.1.131 e evm.model.KI545894.1.122, respectivamente. Outros genes anotados na mapa da Via de Interconversão de Pentose e Glucuronato (KEGG: map00040) estão apresentados na tabela adicional 4.

Uma das formas de analisar o ganho na anotação estrutural é verificando quais funções foram incorporadas na nova anotação em relação à antiga. Como o presente trabalho visa principalmente a exploração de CAZymes e a TAPs, foi verificado em quantas proteínas houve atribuição de funções em cada uma dessas

famílias (tabela adicional 3). É importante também mencionar que alguns destes novos genes foram considerados diferencialmente expressos nos experimentos de “fontes de carbono” (abaixo), enfatizando a importância do melhoramento da anotação da forma como foi feita para a compreensão de questões biológicas, como o metabolismo de diferentes açúcares por *K. brasiliensis* GHG001.

As análises de transcriptoma de *K. brasiliensis* GHG001 em diferentes fontes de carbono

Para a melhor compreensão do metabolismo de pentoses foram realizados experimentos de crescimento de *K. brasiliensis* GHG001 em xilano e xilose, além de glicose, seguido do sequenciamento do transcriptoma usando RNA-seq (Tabela 2). A análise do metabolismo nas diferentes fontes de carbono foi realizada através de contrastes pareados (xilose versus glicose, xilano versus glicose e xilose versus xilano). Os genes considerados diferencialmente expressos foram obtidos a partir da intersecção do resultado de diferentes métodos (Figura 5). Os contrastes revelaram 212 genes positivamente regulados em xilano e 46 em xilose, quando comparados com glicose. Foram considerados negativamente regulados 244 e 25 genes, em xilano e xilose, respectivamente. Quando xilose versus xilano foram comparadas, 182 genes foram positivamente e 205 negativamente regulados.

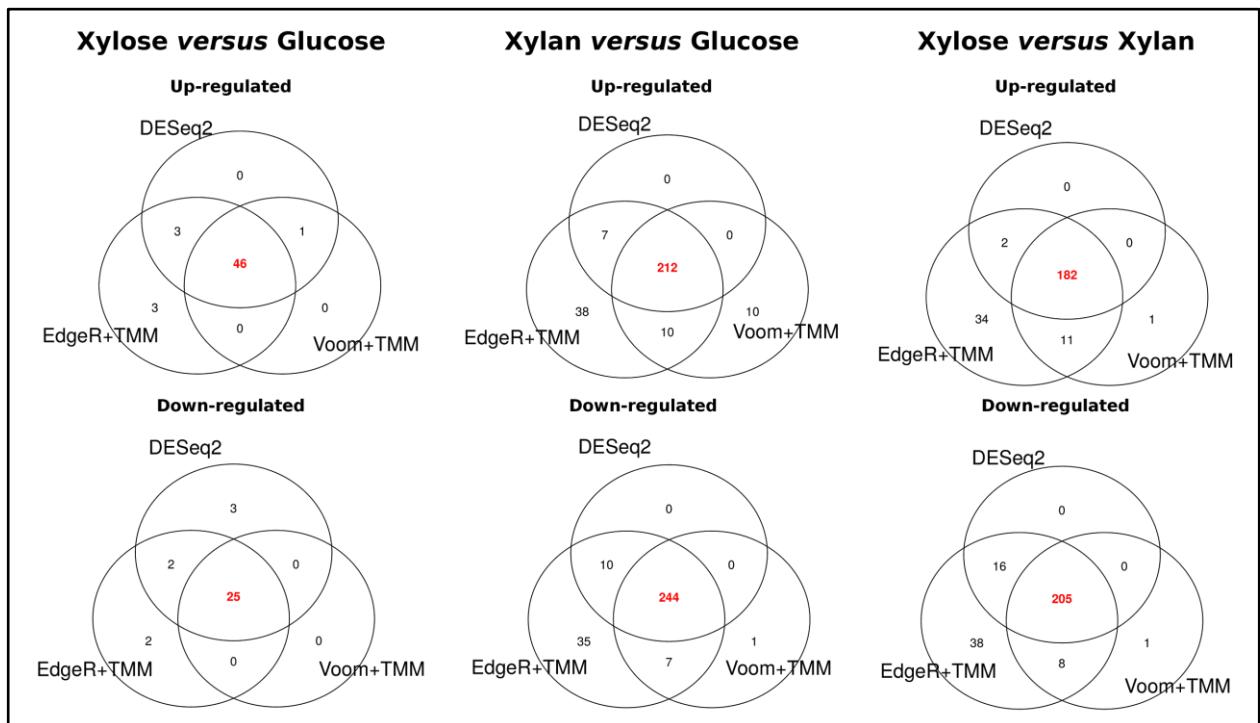


Figura 5 - Diagramas de Venn com genes considerados diferencialmente expressos (DEG) em cada um dos métodos (DESeq2; EdgeR + TMM; Voom + TMM). Em vermelho, o número de genes considerados DE na intersecção entre os métodos considerados, os quais foram usados nas análises subsequentes de enriquecimento de genes e análises de perfil de expressão.

Curiosamente, um número baixo de genes foi considerado diferencialmente expresso nas comparações entre monômeros (xilose ou glicose) e xilano, podendo indicar que uma diversidade maior de mecanismos celulares está envolvida na degradação de polímeros e importação de oligômeros/monômeros.

Entre os genes encontrados exclusivamente na nova anotação (obtida neste trabalho), alguns foram encontrados diferencialmente expressos (Figura 6), entre eles genes que codificam proteínas com possível papel de regulador da transcrição, como um membro da família C2H2 e Fungal Zn(2)-Cys(6). Estas descobertas reforçam a importância do melhoramento da anotação para analisar funções novas e estudar questões específicas, como o metabolismo e regulação de pentoses.

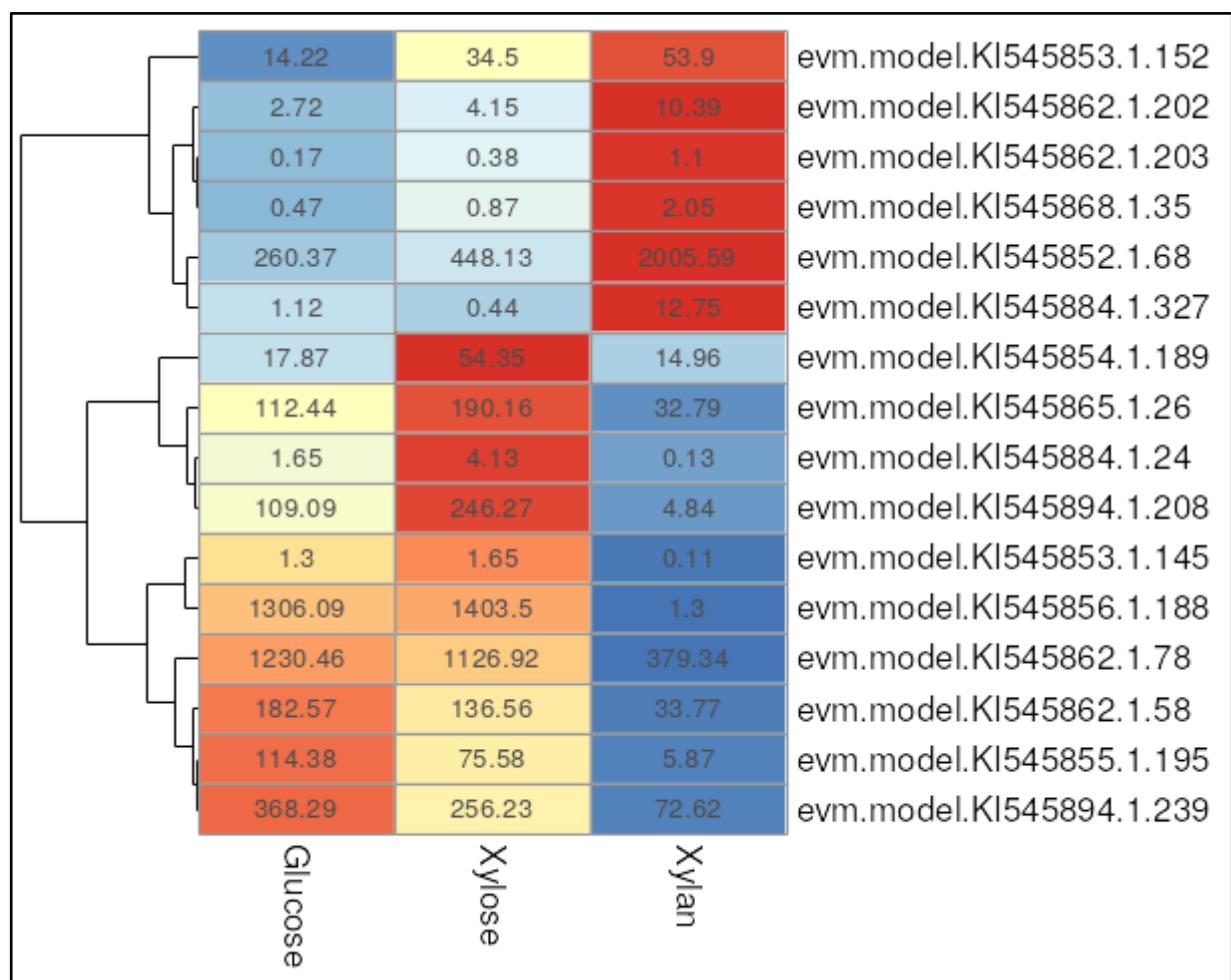


Figura 6 - Genes considerados diferencialmente expressos entre os encontrados exclusivamente na nova anotação (com dados de RNA-seq).

Os genes considerados diferencialmente expressos foram submetidos a uma análise de “clustering” de acordo com o perfil de expressão nas diferentes fontes de carbono, usando Mclust em R. Foram inicialmente gerados 15 “clusters”, os quais foram atribuídos a grupos nomeados de acordo com a indução ou repressão de genes nas diferentes fontes de carbono (xilose, xilano e glicose). Em seguida, os “clusters” com perfis que não concordavam com a descrição do grupo foram realocados aos grupos adequados; quando necessário, genes individuais foram realocados. A tabela 3 e a figura 7 apresenta o resultado final da atribuição de genes diferencialmente expressos aos grupos de perfil de expressão.

Tabela 3 - Número de genes em cada grupo de perfis de expressão (apenas genes diferencialmente expressos).

Grupo	Número de genes
Genes induzidos em xilose (XLOIG)	86
Genes reprimidos em xilose (XLORG)	93
Genes induzidos em xilano (XLAIG)	173
Genes reprimidos em xilano (XLARG)	95
Genes induzidos em glicose (GLUIG)	124
Genes reprimidos em glicose (GLURG)	11

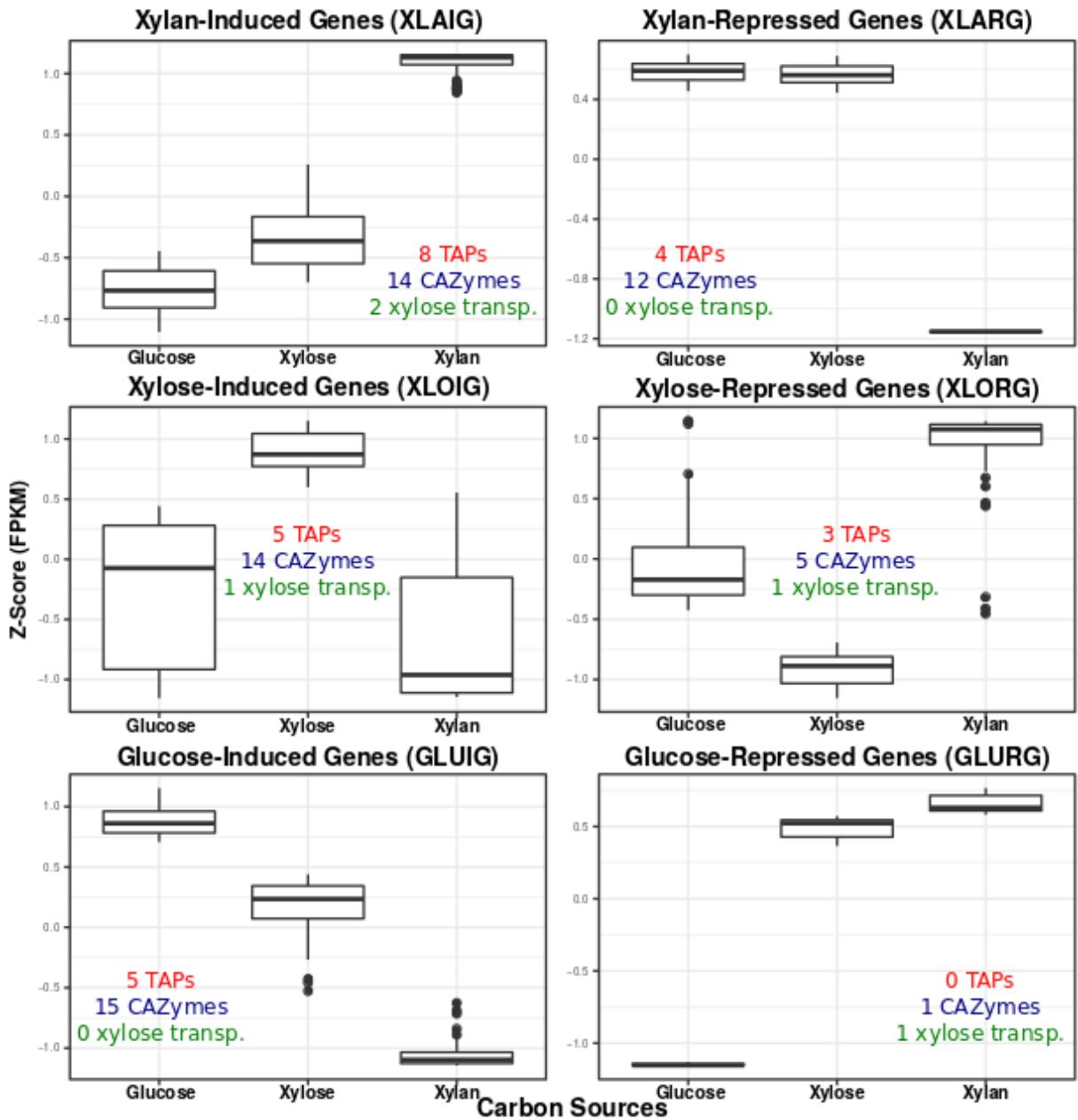


Figura 7 - Análise de perfil de expressão de genes diferencialmente expressos. Genes foram agrupados em seis grupos, de acordo com o perfil de expressão: genes induzidos (XLAIG) e reprimidos (XLARG) em xilano, induzidos (XLOIG) e reprimidos (XLORG) em xilose e genes induzidos (GLUIG) e reprimidos (GLURG) em glicose. O número de genes em cada grupo correspondentes a TAPs, CAZymes e possíveis transportadores de xilose estão destacados em cores vermelho, azul e verde, respectivamente.

Foram analisados os genes envolvidos na degradação de xilano (Dodd e Cann, 2009), os considerados centrais (“core”) na degradação de xilano e xiloglucano em um trabalho recente de transcriptômica comparativa (Peng et al., 2017) e da classe “Auxiliary Enzymes” (AAs), que inclui enzimas ligninolíticas (Levasseur et al., 2013) e mono-oxigenases líticas de polissacarídeos (LPMOs) (Busk e Lange, 2015), as últimas podendo atuar em polissacarídeos da parede

celular de plantas, como hemiceluloses (Agger et al., 2014). A tabela adicional 1 inclui estas enzimas identificadas em *K. brasiliensis* e as possíveis funções associadas a cada família do CAZy Database (<http://www.cazy.org>, Cantarel et al., 2008; Lombard et al., 2013) e os genes diferencialmente expressos estão apresentados na tabela 4.

Como esperado, o gene que codifica a endoxilanase *PbXynA*, previamente descrita por nosso grupo (Borges et al., 2014), aparece diferencialmente expresso e agrupa-se entre os induzidos em xilano (XLAIG). Estudos anteriores em ascomicetos e no basidiomiceto *A. bisporus* (De Groot et al., 1998) mostraram que a endoxilanase é reprimida em glicose, que concorda com o observado neste trabalho. Outras CAZymes possivelmente envolvidas na degradação de xilano apresentam indução no polímero. Entre elas, membros das famílias de GHs: GH3, GH5, GH51 (com CBM4), GH54, GH62 e GH115 (uma proteína da família GH115 com atividade de alfa-glucuronidase apresenta atuação também em cadeias laterais de xilano (Rogowski et al., 2014)), além de uma esterase da família CE1.

Curiosamente, algumas CAZymes apresentam expressão reprimida em xilano (XLARG) ou são induzidas em glicose (GLUIG) ou xirole (XLOIG). Entre elas, enzimas das famílias GH5, GH8, GH43, GH30, GH31 e GH51. Além disso, esterases de carboidratos também foram identificadas entre as famílias de genes induzidos por monômeros. Todos os membros de CE4 diferencialmente expressos e uma das CE1 estão em grupos de enzimas induzidas em xirole ou reprimidas em xilano (ou seja, mais expressas em monômeros). Entre as possíveis funções atribuídas a estas famílias, está a ação em longos polímeros de xilano, facilitando a ação de GHs. Por outro lado, o repertório de CEs em *K. brasiliensis* GHG001 é grande (tabela adicional 1) e entre as possíveis explicações para estas observações estaria: i) a expressão de outras CEs, não diferencialmente expressas, é suficiente para disponibilizar as moléculas de xilano para atuação das GHs, ii) as GHs que atuam nos polímeros de xilano para o substrato utilizado não requerem as CEs para atuarem.

Entre genes que codificam enzimas envolvidas na degradação de lignina (AA1-7), foram identificados diferencialmente expressos nas famílias AA 3, 5 e 7. Curiosamente, membros das famílias 5 e 7 são induzidos apenas em glicose. Embora um membro em AA10 tenha sido identificado em *K. brasiliensis*

(evm.model.KI545859.1.2), o gene não é considerado diferencialmente expresso, embora a maior expressão tenha sido observada em xilose.

Tabela 4 - Genes de *K. brasiliensis* GHG001 que codificam proteínas com domínios de CAZymes possivelmente envolvidas no metabolismo de xilano (Dodd e Cann, 2009; Peng et al., 2017) e que estão incluídos entre os diferencialmente expressos.

Domínio	Gene	Grupo
GH5	evm.model.KI545851.1.140	XLARG
	evm.model.KI545853.1.59	XLAIG
	evm.model.KI545863.1.116	GLUIG
	evm.model.KI545864.1.105	XLOIG
	evm.model.KI545873.1.782	XLORG
GH8	evm.model.KI545851.1.611	XLARG
GH43	evm.model.KI545854.1.192	XLARG
GH51	evm.model.KI545851.1.643 (com CBM4)	XLAIG
	evm.model.KI545864.1.162	XLOIG
GH11	evm.model.KI545855.1.194	XLAIG
GH54	evm.model.KI545862.1.33	XLAIG
GH62	evm.model.KI545864.1.182	XLAIG
GH3	evm.model.KI545863.1.93	XLAIG
	evm.model.KI545866.1.28	XLAIG
CE1	evm.model.KI545862.1.305	XLAIG
	evm.model.KI545863.1.8	XLOIG
CE4	evm.model.KI545864.1.139	XLOIG
	evm.model.KI545873.1.279	XLOIG
	evm.model.KI545884.1.282	XLARG
	evm.model.KI545891.1.114	XLARG
GH115	evm.model.KI545863.1.138	XLAIG
GH30	evm.model.KI545853.1.51	XLARG
GH27	evm.model.KI545858.1.137 (com CBM35)	XLORG
GH31	evm.model.KI545862.1.528	XLOIG
	evm.model.KI545891.1.59	GLUIG

AA3	evm.model.KI545852.1.141	XLARG
	evm.model.KI545854.1.126	XLARG
	evm.model.KI545862.1.274	XLOIG
	evm.model.KI545862.1.439	XLORG
AA5	evm.model.KI545873.1.272	GLUIG
	evm.model.KI545891.1.2	GLUIG
AA7	evm.model.KI545894.1.245	GLUIG

Possíveis transportadores de xilose foram identificados como diferencialmente expressos. A tabela 5 apresenta os resultados de busca com hmmsearch e os grupos nos quais os diferencialmente expressos se encontram. Entre eles, um é reprimido em glicose (expresso nas duas outras condições), dois são induzidos em xilano e um induzido em xilose. A identificação destes genes induzidos em componentes de hemicelulose indicam possíveis alvos para a caracterização funcional e, eventualmente, melhoramento de microrganismos que não têm capacidade de transporte e que poderiam ser aplicados industrialmente. Curiosamente, um deles foi reprimido em xilose, podendo indicar que o transportador pode ser utilizado também (ou preferencialmente) para transporte de glicose ou que é um falso positivo.

Tabela 5 - Identificação de transportadores putativos de xilose em *K. brasiliensis* GHG001.

Gene	E-value	Pontuação hmmsearch	DEG	Grupo
evm.model.KI545893.1.261	0	572,8	Sim	GLURG
evm.model.KI545854.1.36	0	569	Não	-
evm.model.KI545863.1.92	0	544,8	Sim	XLAIG
evm.model.KI545894.1.9	0	506,4	Sim	XLOIG
evm.model.KI545858.1.163	0	501,5	Sim	XLORG
evm.model.KI545873.1.699	0	480,9	Não	-
evm.model.KI545895.1.21	0	477,2	Sim	XLAIG

Em seguida, foram buscados possíveis reguladores. Entre os genes diferencialmente expressos, 28 codificam possíveis TAPs. Entre elas, estão incluídas as famílias NotFungal-ZnClus (quatro membros), CCAAT, GNAT e HMG (três), bHLH (dois) e as famílias com bZIP, C2C2-YABBY, C2H2, FHA, Fork_head, Fungal-trans, Fungal-trans2, HB, LIM e SNF2, apresentando um membro cada. No entanto, embora tenham sido anotadas neste trabalho, pouco se sabe sobre a regulação de

enzimas de degradação de xilanos por TAPs, em basidiomicetos; desta forma, inicialmente foram buscados por ortólogos de reguladores previamente caracterizados.

Recentemente, o trabalho de Todd et al. (2014) buscou ortólogos de diversos reguladores caracterizados em fungos com genoma sequenciados. Os reguladores analisados no trabalho foram buscados através de ortologia entre *M. maydis* e *K. brasiliensis* GHG001 (tabela 6). Entre eles, o ortólogo de *hapE* e de *alcR* em *A. nidulans* foram detectados entre os genes diferencialmente expressos. O *hapE* é conhecido como um regulador positivo, ativador de transcrição, enquanto *alcR* é conhecido como regulador da utilização de etanol em *A. nidulans* (Todd et al., 2014). Entre os reguladores de CAZymes mais estudados em fungos, o homólogo do regulador *CreA* (evm.model.KI545854.1.176), embora não seja considerado diferencialmente expresso nos contrastes realizados, apresenta alta expressão em número de “reads” (ou FPKM) nas três condições estudadas. Não foram encontrados homólogos para os reguladores *XlnR* e *Ace2*. Embora homólogos de *HapB* e *HapE* tenham sido encontrados, apenas *HapE* (evm.model.KI545862.1.477) foi considerado diferencialmente expresso (GLUIG).

Tabela 6 - Genes de *K. brasiliensis* GHG001 que codificam proteínas com ortólogos em reguladores caracterizados (Todd et al., 2014).

Gene	Família (Todd et al)	Identificador regulador caracterizado	Homólogo em <i>K. brasiliensis</i> GHG001	Família TAPs	DEG	Grupo
metR	bZIP	AN4361	evm.model.KI545873.1.749	bZIP	Não	-
con7	C2H2	MGG_05287	evm.model.KI545891.1.84	Não identificada	Não	-
hapB	CBF	AN7545	evm.model.KI545862.1.151	CCAAT	Não	-
hapE	CBF	AN6492	evm.model.KI545862.1.477	CCAAT	Sim	GLUIG
creA	FOG Zn finger	AN6195	evm.model.KI545854.1.176	C2H2	Não	-
alcR	GAL4	AN8978	evm.model.KI545852.1.156	Não identificada	Sim	XLORG
uaY	GAL4-Fungal-trans	AN0891	evm.model.KI545873.1.503	Fungal-ZnClus	Não	-

WC2	GATA	Schco1.fgenes h1_pm.C_scaff old_2000088	evm.model.KI5458 91.1.142	C2C2-GATA	Não	-
dopA	Leucine zipper	AN2094	evm.model.KI5458 55.1.167	Não identificada	Não	-
rImA	SRF-type	AN2984	evm.model.KI5458 64.1.51	MADS	Não	-
medA	unknown(med A)	AN6230	evm.model.KI5458 52.1.188	Não identificada	Não	-

Recentemente, Kaupert Neto et al. (2015) buscaram possíveis enzimas envolvidas no metabolismo de hemiceluloses no secretoma de *K. brasiliensis* GHG001 usando dados de espectrometria de massas (análise proteômica, cujos dados que foram incorporados durante o melhoramento da anotação estrutural apresentada neste trabalho). Diversas CAZymes foram detectadas em fontes de carbono incluindo xilose, xilano e glicose, analisados neste trabalho. A ausência e a presença de genes nos secretomas, bem como a anotação funcional e grupo de genes nas análises de expressão diferencial do presente trabalho estão apresentados na tabela 7.

Algumas CAZymes secretadas foram também detectadas entre outras possivelmente envolvidas na degradação de xilano e reveladas nas análises de RNA-seq (tabela 4). Os resultados de Kaupert Neto et al. (2015) corroboram claramente alguns dos resultados. Por exemplo, para as enzimas codificadas por evm.model.KI545891.1.114 (XLARG), evm.model.KI545866.1.28 (XLAIG), evm.model.KI545863.1.93 (XLAIG), evm.model.KI545864.1.182 (XLAIG), foram reportadas a ausência e a presença no secretoma induzido por xilano que correspondem a baixa ou a alta expressão de genes, respectivamente. Em especial, as duas GH3 evm.model.KI545866.1.28 e evm.model.KI545863.1.93 foram validadas usando PCR em tempo real.

Já em outros casos, no entanto, como do gene de GH5 evm.model.KI545873.1.782 (XLORG), a proteína é detectada no secretoma induzido por xilose, indicando tanto a discrepância/ diferença em sensibilidades entre as técnicas usadas, como na ausência de correlação necessária entre a produção de transcritos e a geração de produto final (proteína secretada). Certa discrepância é observada também para a endoxilanase *PbXynA* (XLAIG) - a proteína foi encontrada no secretoma nas três fontes (xilose, xilano e glicose) mas o gene foi induzido em

xilano nas análises de expressão gênica. Isso pode indicar que embora haja a baixa expressão do gene em glicose quando comparada com outras condições experimentais, a evolução do organismo pode ter sido direcionada para otimização da produção e exportação da proteína, que continua sendo observada em glicose.

Tabela 7 - Genes identificados no secretoma de *K. brasiliensis* GHG001 em crescimento em diferentes fontes de carbono (xilose, xilano e glicose) e expressão no presente trabalho.

Identificador no Genbank	Identificador no proteoma melhorado	Secretoma			Anotação funcional *	DEG	Grupo (DEG)
		Xilose	Xilano	Glicose			
EST04645.1	evm.model.KI545895.1.203	X		X	GH15	Não	-
EST04759.1	evm.model.KI545894.1.117			X	AA7	Não	-
EST04859.1	evm.model.KI545894.1.245	X	X	X	AA7	Sim	GLUIG
EST05464.1	evm.model.KI545891.1.79	X			GH25	Não	-
EST05499.1	evm.model.KI545891.1.114	X		X	CE4	Sim	XLARG
EST06571.1	evm.model.KI545873.1.615	X		X	CE4	Não	-
EST06726.1	evm.model.KI545873.1.782	X		X	GH5	Sim	XLORG
EST06841.1	evm.model.KI545868.1.113	X			GH128	Não	-
EST06953.1	evm.model.KI545866.1.28			X	GH3	Sim	XLAIG
EST07255.1	evm.model.KI545864.1.182	X	X		GH62	Sim	XLAIG
EST07368.1	evm.model.KI545863.1.93			X	GH3	Sim	XLAIG
EST08778.1	evm.model.KI545856.1.37	X	X	X	AA7	Não	-
EST09088.1	evm.model.KI545855.1.194 **	X	X	X	GH11	Sim	XLAIG
EST09859.1	evm.model.KI545851.1.252	X		X	-	Não	-
EST09945.1	evm.model.KI545851.1.344	X		X	-	Não	-
EST10058.1	evm.model.KI545851.1.464	X	X	X	-	Não	-
EST10227.1	evm.model.KI545851.1.643		X		CBM4, GH51	Sim	XLAIG
EST10106.1	evm.model.KI545851.1.516	X	X	X	-	Não	-
EST10013.1	evm.model.KI545851.1.414	X		X	-	Sim	XLAIG
EST09784.1	evm.model.KI545851.1.168	X	X	X	-	Não	-
EST09697.1	evm.model.KI545851.1.72	X	X	X	-	Sim	XLORG
EST09179.1	evm.model.KI545854.1.89	X		X	-	Sim	XLARG
EST08904.1	evm.model.KI545856.1.187	X			-	Sim	XLARG
EST08506.1	evm.model.KI545858.1.104	X	X	X	-	Sim	XLAIG
EST08492.1	evm.model.KI545858.1.89	X		X	-	Não	-
EST08405.1	evm.model.KI545859.1.164	X	X	X	-	Sim	XLARG

EST08037.1	evm.model.KI545861.1.86	X	X	X	-	Sim	XLAIG
EST07270.1	evm.model.KI545864.1.197		X		-	Sim	XLORG
EST06764.1	evm.model.KI545868.1.19	X		X	-	Não	-
EST06419.1	evm.model.KI545873.1.449	X	X		-	Não	-
EST06201.1	evm.model.KI545873.1.214			X	-	Sim	XLAIG
EST06015.1	evm.model.KI545873.1.2	X		X	-	Não	-
EST05834.1	evm.model.KI545884.1.197			X	-	Não	-
EST05473.1	evm.model.KI545891.1.88			X	-	Sim	XLAIG
EST05303.1	evm.model.KI545892.1.199	X			-	Não	-
EST05061.1	evm.model.KI545893.1.197			X	-	Não	-
EST05050.1	evm.model.KI545893.1.186	X		X	-	Sim	XLARG
EST04518.1	evm.model.KI545895.1.60	X	X	X	-	Não	-

* Fonte de carbono na qual a proteína codificada pelo gene foi encontrada (Kaupert Neto et al., 2014).

** *PbXynA*, endoxilanase caracterizada em Borges et al. (2014).

Por fim, foi analisada a expressão de alguns genes considerados importantes para o metabolismo interno de xilose. Entre eles, estão os associados ao mapa das vias de Interconversão de Pentose e Glucuronato (map00040). O gene que codifica a XDH (evm.model.KI545894.1.122) foi considerado diferencialmente expresso, sendo expresso principalmente em xilose e xilano (GLURG). Além dele, o gene evm.model.KI545891.1.306 está entre os genes induzidos em xilano.

Curiosamente, o gene de XR (evm.model.KI545895.1.131) não foi considerado diferencialmente expresso. Assim, embora não tenham sido considerados diferencialmente expressos, a inspeção manual dos valores de expressão para alguns genes de interesse e de conhecimento prévio sobre envolvimento no metabolismo de xilose sugere que muitos genes podem não terem sido considerados diferencialmente expressos devido à estrengência dos métodos aplicados.

Para ajudar na interpretação dos resultados, análises de enriquecimento de mapas do KEGG, famílias de CAZymes e TAPs foram realizadas para cada um dos grupos de perfis de expressão (Tabela 8; Figura 7). Foram consideradas significativamente enriquecidas classes (mapas ou famílias) com p valor corrigido < 0,05. O maior número de mapas do KEGG enriquecidos foi observado em grupos de genes induzidos (três) ou reprimidos (quatro) em xilose. Entre os mapas enriquecidos no grupo de genes induzidos com xilose (XLOIG), estão “map00051 -

Metabolismo de Frutose e Manose”, “map00520 - Amino sugar and nucleotide sugar metabolism” e “map00030 - Via de Pentoses Fosfato”. Em genes reprimidos (XLORG), aparecem “map01120 - Metabolismo microbiano em ambientes diversos”, “map00190 - Oxidação fosforilativa”, “map02020 - Sistema de Dois Componentes” e “map00020 - Ciclo do Ácido Cítrico”. Apenas um mapa foi considerado enriquecido para os genes reprimidos em glicose (“map00480 - Metabolismo de Glutathione”) e em genes induzidos em xilano (“map03030 - Replicação de DNA”). Uma anotação detalhada destes genes é apresentada na tabela adicional 4. Nenhuma família de CAZyme foi considerada enriquecida nas condições estudadas.

Tabela 8 - Análise de enriquecimento de mapas do KEGG e famílias de CAZymes e TAPs para os grupos de genes diferencialmente expressos.

Grupo	Análise	Classe enriquecida *	P valor	Odds Ratio	P valor corrigido
GLURG	KEGG	map00480 (Glutathione metabolism)	0,004	28,894	0,048
XLAIG	KEGG	map03030 (DNA replication)	0,0001	9,944	0,011
XLOIG	KEGG	map00051 (Fructose and Mannose Metabolism)	0,0003	15,008	0,013
XLOIG	KEGG	map00520 (Amino Sugar and Nucleotide Sugar Metabolism)	0,0006	8,704	0,013
XLOIG	KEGG	map00030 (Pentose Phosphate Pathway)	0,0008	11,755	0,013
XLORG	KEGG	map01120 (Microbial metabolism in diverse environments)	0,006	3,936	0,049
XLORG	KEGG	map00190 (Oxidative phosphorylation)	0,0001	9,465	0,002
XLORG	KEGG	Map02020 (Two-Component System)	0,00001	40,403	0,0004
XLORG	KEGG	Map00020 (Citrate cycle (TCA cycle))	0,0001	20,333	0,002
XLORG	TAPs	NotFungal-ZnClus	0,0207	22,019	0,041

* A análise de enriquecimento é realizada com diferentes classes. Nas análises do KEGG, são analisados os mapas; nas análises de CAZymes e TAPs, são analisadas famílias.

Para alguns dos genes associados aos mapas enriquecidos, foram identificados domínios de CAZymes no produto proteico, como a GH51 codificada por evm.model.KI545864.1.162, associada ao mapa map00520, e induzida em xilose.

Entre as vias identificadas nas análises de enriquecimento de mapas do KEGG, destaca-se a Via de Pentoses Fosfato (PPP). Diversos trabalhos destacaram a necessidade de superexpressão de genes desta via em *Saccharomyces cerevisiae*, para melhorar a fermentação de xilose (Matsushika et al., 2012). A maior parte dos genes identificados no presente trabalho e atribuídos a esta via está agrupada entre os genes induzidos em xilose (XLOIG) ou entre os genes reprimidos em glicose (ainda mais expressos em xilose). Entre os genes, foi identificado o gene evm.model.KI545866.1.6 que codifica uma transaldolase, considerada chave para o crescimento em xilose e consumo de xilulose em *S. cerevisiae* (Matsushika et al., 2012).

A identificação de PPP como enriquecida em XLOIG e, em especial, de genes previamente estudados e cuja expressão é considerada importante para uso de xilose e consumo de xilulose, como é o caso da transaldolase, abre boas perspectivas para a exploração em laboratório de genes em outras vias consideradas enriquecidas. Por exemplo, seria interessante estudar com maior profundidade genes que foram reprimidos em xilose, como do TCA, Fosforilação Oxidativa e de Sistema de Dois Componentes. Seria interessante, por exemplo, subexpressar alguns destes genes em cepas industriais para avaliar o aumento do consumo de xilose.

CONCLUSÕES

- A incorporação de dados experimentais de transcriptômica usando RNA-seq melhorou de forma substancial a anotação estrutural de *K. brasiliensis* GHG001, provendo uma plataforma acurada para as análises de transcriptoma direcionadas a perguntas específicas, como o metabolismo de açúcares, e para análises de genômica comparativa.
- O número de genes diferencialmente expressos é maior na comparação envolvendo monômeros e o polímero xilano, quando comparados com os contrastes entre monômeros (*xilose versus glicose*).
- A análise exploratória de CAZymes, particularmente as envolvidas no metabolismo de xilano, e proteínas reguladoras da transcrição (TAPs), permite gerar um maior entendimento acerca do metabolismo de xilano e xilose na levedura.
- As análises de RNA-seq dão evidências sobre a regulação das CAZymes através da indução ou repressão de enzimas nas diferentes fontes de carbono, incluindo enzimas não diretamente envolvidas com o metabolismo de xilanos e suas cadeias laterais, como as possivelmente envolvidas na degradação de lignina.
- As evidências sobre a regulação nas análises de RNA-seq são ainda mais relevantes se forem considerados o baixo número de estudos sobre a regulação transcracional de CAZymes em basidiomicetos, a baixa conservação de homólogos destes reguladores entre os diferentes grupos de fungos e de função, para os casos em que homólogos foram encontrados, devido ao “transcriptional rewiring”; estes fatores impedem a simples transferência de função entre homólogos.
- Os resultados de RNA-seq dão maior evidência para a possível função de proteínas identificadas como possíveis transportadores de xilose. Em particular, são relevantes os casos de genes mais expressos em xilose e xilano, podendo ser alvos de caracterização funcional no futuro. Além disso, destacam-se genes cuja expressão é maior em xilose e xilano, indicando um possível transportador específico de xilose.
- A análise de enriquecimento de mapas do KEGG, CAZymes e TAPs dá indicações de grupos de genes importantes a serem explorados em trabalhos futuros. Em particular, mapas do KEGG sugerem exploração de genes

envolvidos na replicação de DNA, Ciclo do Ácido Cítrico e Fosforilação Oxidativa, Via de pentoses fosfatos e do metabolismo de açúcares, como frutose e manose. Parece evidenciada também a importância de Sistemas de Dois Componentes, importante no sensoriamento do ambiente e início de vias de sinalização.

- Na análise de enriquecimento, a Via de Pentoses Fosfato foi considerada enriquecida com genes induzidos em xilose. Entre os genes diferencialmente expressos nesta via, há uma transaldolase, enzima considerada importante no metabolismo de xilulose e entre as responsáveis pelo direcionamento deste substrato para as vias glicolíticas.

Capítulo 2 - Análises de genômica comparativa de *Ustilaginaceae*, incluindo o sequenciamento de novas leveduras da família

INTRODUÇÃO

A diversidade de leveduras é enorme e inclui ascomicetos e basidiomicetos cujo estado sexual, se conhecido, não é envolvido por um corpo frutificante (Kurtzman et al., 2015). As aplicações industriais e o potencial biotecnológico de leveduras são amplos e as relações filogenéticas, cujo entendimento tem sido melhorado com estudos recentes, refletem esta amplitude/diversidade. Desta forma, embora muitas vias metabólicas sejam compartilhadas entre a maioria das leveduras, como as de fermentação de glicose, outras parecem ser restritas a linhagens filogenéticas específicas, como de utilização de metanol ou fermentação de xilose. Ao mesmo tempo, a produção de alguns compostos parece estar correlacionada com a filogenia, como é o caso da produção de micosporinas por leveduras basidiomicéticas (Kurtzman et al., 2015).

A levedura ascomicética *S. cerevisiae* é amplamente utilizada na produção de pães, cerveja e vinho e genomas de diferentes cepas da espécie têm sido amplamente explorados por diversos grupos de pesquisa (Mewes et al., 1997; Engel et al., 2014). No entanto, apesar da genômica de *S. cerevisiae* ter esclarecido muito sobre a biologia de eucariotos, esta levedura apresenta alta divergência fisiológica com outras leveduras e a genômica comparativa incluindo organismos pouco explorados pode revelar o potencial biotecnológico dos mesmos, incluindo espécies capazes de crescer em metanol como única fonte de carbono e energia, que produzem altas quantidade de vitaminas e lipídios ou que fermentam fontes de carbono não convencionais (Riley et al., 2016). No entanto, a maior parte dos trabalhos têm focado em ascomicetos, sendo que poucos trabalhos envolvem a exploração de basidiomicetos (Hittinger et al., 2015; Riley et al., 2016).

Os basidiomicetos constituem os mais eficientes degradadores de biomassa e, como as enzimas codificadas pelo genoma têm grande chance de refletir a adaptação a nichos específicos, estes fungos têm um grande potencial para a exploração na indústria. Diversos basidiomicetos, isolados de uma variedade de nichos e vivendo em substratos variados, já foram sequenciados e a exploração de enzimas de metabolismo de carboidratos permite fazer uma ligação com seus

habitats naturais (Rytioja et al., 2014), sendo geralmente categorizados de acordo com a sua ecologia, que pode ser associada com a distribuição de CAZymes, como a degradação branca ou branda, decompositores de palha e resíduos de lixo, ectomicorrízicos simbiontes, fungos fitopatogênicos ou micoparasitas e leveduras.

A genômica comparativa permite comparar o tamanho e a arquitetura de genomas, além do conteúdo e a estrutura de genes (Miller et al., 2004). Em especial, o sequenciamento de genomas de fungos tem sido cada vez mais realizado e revisões recentes já exploraram avanços na área (Sharma, 2016). Os projetos geralmente visam entender relações entre interações de fungos e plantas, obter novos “insights” sobre a degradação de biopolímeros e explorar o potencial ainda não descoberto de arsenais naturais com aplicação potencial (Grigoriev et al., 2013). Diversos estudos de genômica comparativa foram realizados em fungos, incluindo análises de degradadores de biomassa (Nagy et al., 2015; Sharma et al., 2015; Riley et al., 2014; Ohm et al., 2014; Benoit et al., 2015).

Em especial, métodos comparativos filogenéticos (PCM) envolvem o estudo de mecanismos de origem de diversidade em um contexto de história evolutiva. Muitos destes métodos são revisados no livro de Garamszegi (2014) e uma filogenia precisa do grupo em estudo é esperada para dar suporte aos métodos de PCM. Para a inferência filogenética, métodos de filogenômica, que incluem o uso de uma supermatriz ou de abordagens “supertree”, têm sido explorados nos últimos anos devido ao aumento na disponibilidade de genomas completos para diversos grupos de organismos (Delsuc et al., 2005). Há aproximadamente dez anos, os primeiros trabalhos usando informação de genoma completo foram realizados para a reconstrução da história evolutiva em fungos através de filogenômica. Exemplos incluem as análises de Fitzpatrick et al. (2006), Kuramae et al. (2006) e Cornell et al. (2007). Trabalhos mais recentes incluem Ebersberger et al., (2011) e Medina et al. (2011). Métodos ainda mais recentes usando a abordagem filogenômica focaram na seleção de alguns marcadores informativos, permitindo assim uma análise de novos fungos com apenas alguns poucos alinhamentos (Capella-Gutierrez et al., 2014).

Neste trabalho, são analisadas contrações e expansões de famílias de CAZymes e TAPs em *Ustilaginaceae* usando o software CAFE v4 (Hahn et al., 2005; Han et al., 2013), que assume um processo de nascimento e morte para a evolução de famílias gênicas. O processo, originado em campos da matemática e originalmente desenvolvido para estudos de crescimento populacional incorporando

fatores estocásticos, tem diversas aplicações na biologia moderna, como na estimativa de idade de alelos, reconstrução filogenética e modelagem de diversos aspectos da evolução de genomas (Novozhilov et al., 2006). Adicionalmente, o CAFE requer uma precisa filogenia do grupo em estudo com tempo (cronograma), que prevê um modelo nulo para teste de hipóteses (se uma família teve expansão ou contração significativa). Diversos métodos recentes usando abordagem Bayesiana têm se destacado para a estimativa de idade dos nós em filogenias, tendo datações fósseis para alguns organismos (como os nós ancestrais) incluídos nas análises. Em particular, a importância de métodos Bayesianos inclui a possibilidade de incorporar incertezas nas datações fósseis e também variações de taxa de evolução em diferentes ramos das árvores filogenéticas, para grupos de organismos próximos do ponto de vista evolutivo/ filogenético (Dos Reis et al., 2016).

Diversos trabalhos foram realizados usando o CAFE para estudar a evolução de famílias gênicas em fungos, incluindo de genes relacionados à degradação de biomassa. Floudas et al. (2012) estudaram a evolução de desconstrução de lignina em fungos em Agaricomycetes usando genômica comparativa, incluindo a evolução do tamanho de famílias de peroxidases (PODs). Recentemente, bases genéticas de hábitos endofíticos foram estudados usando análises de expansão e contração de famílias de CAZymes em fungos ascomicetos, com ênfase em Xylonomycetes (Gazis et al., 2016) e na adaptação a estilos de vida em diferentes relações com hospedeiros em patógenos de plantas (Gan et al., 2016).

Recentemente, Berlemont (2017) analisou a distribuição e a diversidade da degradação de polissacarídeos em fungos usando uma abordagem genômica, permitindo analisar o potencial em todos os genomas disponíveis no Mycocosm (Grigoriev et al., 2013). Eles enfatizaram a alta frequência de celulases em Ustilaginomycetes; no entanto, incluíram apenas seis de seus membros e apenas quatro da família *Ustilaginaceae* (*M. maydis* 521, *Moesziomyces antarcticus* T-34, *Pseudozyma hubeiensis* SY62 and *S. reilianum* SRZ2).

Em *Ustilaginaceae*, tendo em mente a importância de alguns fungos de carvão, como *M. maydis*, *S. scitamineum* e *S. reilianum* e de alguns organismos com forma exclusivamente assexual (anamórficos) e a disponibilidade de genomas completos, um estudo foi realizado para estudar a evolução da patogenicidade, sugerindo que *A. flocculosa* teria sido resultado da transição de fungos patogênicos para uma forma anamórfica não infecciosa (Lefebvre et al., 2013). Embora tenha

incluído a anotação de CAZymes, eles claramente focaram em enzimas envolvidas na degradação de parede celular de fungos (*A. flocculosa* é um patógeno de fungos e agente de biocontrole), não tendo análises específicas de enzimas de degradação de parede celular de plantas; além disso, eles não estudaram a evolução de tamanhos de famílias de CAZymes, tampouco de reguladores de transcrição. Por fim, o estudo incluiu poucos organismos, sendo *A. flocculosa* a única anamórfica da família de fungos.

Além das CAZymes, trabalhos anteriores envolvendo a genômica comparativa exploraram também FTs em eucariotos, incluindo fungos. Famílias foram reportadas como enriquecidas, com ganhos e perdas ao longo da filogenia. Exemplos de famílias que apresentaram ganhos em fungos são MAT-alpha, Copper_fist e fungal_TF2, enquanto STAT e RHD foram perdidas e NDT80 e Zn_clus apresentaram enriquecimento (de Mendoza et al., 2013). Estudos subsequentes analisaram reguladores em fungos, como o trabalho de Todd et al. (2014). Eles focaram particularmente na busca por 64 reguladores transpcionais previamente caracterizados em ascomicetos e basidiomicetos. Os reguladores foram categorizados em diferentes grupos, de acordo com o envolvimento no desenvolvimento e morfologia, metabolismo de carboidratos e metabolismo de aminoácidos e nitrogênio. Mais recentemente, Shelest (2017) descreveu uma visão geral de fatores de transcrição em fungos, chamando-os de “TFomes”. Comparado com Todd et al. (2014), o trabalho realizou uma análise com maior amostragem de taxa de fungos, incluindo alguns membros da *Ustilaginaceae*, como *M. maydis* 521, *M. antarcticus* T-34, *P. hubeiensis* SY62, e *S. reilianum* SRZ2. Embora os trabalhos desenvolvidos até o momento tenham explorado CAZymes e reguladores em fungos, nenhum foca especificamente na família *Ustilaginaceae* de basidiomicetos.

As brácteas de flores de *Heliconia* (Heliconiaceae) formam um microambiente que apresenta diversos insetos e microrganismos, sendo o néctar destas flores a fonte de alimento para estes organismos (Schnittler e Stephenson, 2002; Barbosa et al., 2012; Landell et al., 2014). Tendo em vista o potencial biotecnológico de *K. brasiliensis* GHG001 e de outros membros da família *Ustilaginaceae*, para o desenvolvimento do presente trabalho foram selecionadas leveduras inicialmente identificadas como pertencentes ao grupo polifilético “*Pseudozymas*”, cujos membros estão atribuídos a diferentes novos gêneros (Wang et al., 2015), isoladas de flores de *H. psittacorum*, para sequenciamento e inclusão em análises de genômica

comparativa. As análises apresentadas neste capítulo visam explorar o genoma de *Ustilaginaceae*, em especial das novas leveduras sequenciadas, com foco nas famílias enzimas do metabolismo de carboidratos e em possíveis reguladores da transcrição. Além disso, os resultados de evolução de famílias gênicas em *K. brasiliensis* GHG001 permite uma melhor compreensão do metabolismo de xilano e xilose nesta levedura, tendo como “background” os trabalhos previamente publicados explorando o organismo e as análises de RNA-seq apresentadas no capítulo 1 desta dissertação.

MATERIAIS E MÉTODOS

Sequenciamento de novas leveduras da família *Ustilaginaceae*

Diversos isolados de levedura foram obtidos de flores de *H. psittacorum* disponibilizados pela coleção de cultura da Universidade Federal de Minas Gerais (UFMG, Dr. Carlos Augusto Rosa), previamente obtidos por Barbosa (2010). Os seguintes isolados foram sequenciados no presente estudo e incluídos nas análises comparativas: *Moesziomyces spp.* F16C1 e F5C1 e *Pseudozyma sp.* F8B2 pro tempore. Detalhes sobre a seleção das leveduras do estudo são descritos na seção de resultados. Para extração de DNA, foi feito um pré-inóculo em meio YPD, a 30°C, 200 rpm. Após 24 horas de crescimento, o “pellet” da levedura foi coletado e a extração de DNA genômico foi feita com o kit “DNeasy Plant Mini” (Qiagen, <https://www.qiagen.com/us/shop/sample-technologies/dna/genomic-dna/dneasy-plant-mini-kit/#orderinginformation>). A lise celular é feita com pérolas de vidro em *vortex* durante dez minutos. Os genomas foram sequenciados no USC Epigenome Center (Los Angeles, Califórnia, EUA), com biblioteca TruSeq. A qualidade do sequenciamento foi verificada com FastQC v.0.11.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) e limpos usando Trimmomatic (Bolger et al. 2014). Busca com BLAST contra o banco de nucleotídeos do NCBI (nt) (Altschul et al., 1990), seguida do uso do programa MEGAN5 (Huson e Mitra 2012) foram realizados para verificar a presença de possíveis contaminantes. KmerGenie (Chikhi e Medvedev, 2014) foi usado para estimar o melhor tamanho de k-mer para montagens usando grafos de Bruijn e a média de comprimento do inserto foi calculada com SAMtools (Li et al., 2009), após alinhamento com Bowtie (Langmead e Salzberg, 2012). A comparação dos diferentes montadores, SPAdes

v.2.4 (Bankevich, Nurk et al. 2012), IDBA-UD (Peng, Leung et al. 2012), Velvet v.1.1.06 (Zerbino and Birney 2008) e SAGE (Ilie, Haider et al. 2014), foi realizada usando estatísticas de montagens geradas pelo QUAST v.2.1 (Gurevich et al., 2013) e CEGMA (Parra et al., 2007). As montagens geradas pelo programa SAGE foram selecionadas para análises subsequentes. O melhoramento destas montagens foi realizado com IMAGE (Tsai et al. 2010) e iCORN (Otto et al., 2010), incluídos no pipeline PAGIT (Swain et al., 2012), seguido dos programas SSPACE Basic v 2.0 (Boetzer et al., 2010), GapFiller v. 1.11 (Nadalin e Policriti, 2012) e o “pipeline” Pilon (Walker et al., 2014). Análises de completeza dos genomas foram realizadas usando BUSCO (Simão et al., 2015) com um conjunto de 1.438 ortólogos presentes em fungos. Dados fenotípicos foram obtidos como descritos em de Castro Oliveira et al. (2014). A comparação da região ITS e ribossomal entre as espécies filogeneticamente relacionadas com as sequenciadas neste projeto foi realizada usando alinhamentos pareados com needle, do pacote EMBOSS 6.6.0.0 (Rice et al., 2000).

Genomas usados na análise comparativa

Genomas usados nas análises comparativas estão apresentados na tabela 9. Estatísticas gerais dos genomas são apresentadas na seção de resultados deste capítulo.

Tabela 9 - Genomas de Ustilaginomycotina incluídos nas análises de filogenômica e genômica comparativa (basidiomicetos incluídos na figura 9 estão apresentados na tabela adicional 5).

Identificador	Organismo	Banco de dados	Identificador no banco de dados	Referência	# CAZymes	# TAPs
AFL1	<i>Anthracocystis flocculosa</i> PF-1	NCBI	GCA_000417875.1	Lefebvre et al. (2013)	272	319
KBR1	<i>K. brasiliensis</i> GHG001	NCBI	GCA_000497045.1	de Castro Oliveira et al. (2013)	262	296
MAN1	<i>Moesziomyces antarcticus</i> JCM10317	NCBI	GCA_000747765.1	Saika et al. (2014)	289	319
MAN2	<i>Moesziomyces antarcticus</i> T-34	NCBI	GCA_000334475.1	Morita et al. (2013)	282	325
MAP1	<i>Moesziomyces aphidis</i> DSM70725	NCBI	GCA_000517465.1	Lorenz et al. (2014)	264	277
MF16	<i>Moesziomyces</i> sp. F16C1	Presente trabalho	Presente trabalho.	Presente trabalho.	293	325

MF51	<i>Moesziomyces sp.</i> <i>F5C1</i>	Presente trabalho	Presente trabalho.	Presente trabalho.	288	323
PHU1	<i>Pseudozyma</i> <i>hubeiensis SY62</i>	NCBI	GCA_000403515.1	Konishi et al. (2013)	281	313
MF82	<i>Pseudozyma sp.</i> <i>F8B2</i>	Presente trabalho	Presente trabalho	Presente trabalho	284	289
SRM1	<i>Sporisorium</i> <i>reilianum SRZ2</i>	NCBI	GCA_000230245.1	Schirawski et al. (2010)	265	316
SSC2	<i>Sporisorium</i> <i>scitamineum</i> <i>2014001</i>	NCBI	GCA_000772675.1	Que et al. (2014)	255	309
SSC1	<i>Sporisorium</i> <i>scitamineum</i> <i>CSIRO</i>	NCBI	GCA_001243155.1	-	237	302
SSC3	<i>Sporisorium</i> <i>scitamineum</i> <i>SSC39</i>	NCBI	GCA_001010845.1	Taniguti et al. (2015)	254	302
UES1	<i>Ustilago esculenta</i> <i>MMT</i>	NCBI	GCA_000819925.1	Ye et al. (2017)	250	307
UHO1	<i>Ustilago hordei</i> <i>Uh4857-4</i>	NCBI	GCA_000286035.1	Laurie et al. (2012)	241	303
UMA1	<i>Mycosarcoma</i> <i>maydis 521</i>	NCBI	GCA_000328475.1	Kämper et al. (2006)	274	318
MGL1	<i>Malassezia</i> <i>globosa</i> *	JGI	Malgl1	Xu et al. (2007)	103	161

* Grupo externo.

** Os domínios são contados independentemente, isto é, se uma mesma proteína apresenta dois domínios diferentes, é considerada uma soma de dois.

Identificação de ortólogos nos proteomas

A identificação de proteínas ortólogas foi realizada com BLAST (Altschul et al., 1990) e OrthoMCL v. 2.0.9 (Li et al., 2003), cujos grupos de ortólogos foram determinados com valor de inflação de 1,5. Ortólogos de cópia única presentes em todos os taxa, ou em todos com exceção de um, foram extraídos usando script Perl “in house”.

Reconstrução filogenética

Análises filogenômicas foram realizadas para a reconstrução da árvore de espécies para a família *Ustilaginaceae*, usando sequências de aminoácidos de quatro marcadores selecionados por Capella-Gutiérrez et al. (2014) (tabela 10), que possibilitam a discriminação de espécies de fungos. A busca pelos marcadores nos proteomas estudados foi realizada usando o programa HMMER3 (Eddy, 2009) e perfis HMM desenvolvidos a partir de dados fornecidos pelos autores (8.856 proteínas de basidiomicetos, entre as quais as proteínas dos quatro marcadores

estavam marcadas). Para cada HMM criado, a seleção de um limiar de pontuação (GA) foi selecionada como descrito em Perez-Rodríguez et al. (2009). No entanto, neste trabalho foi usado um método iterativo, desenvolvido em script “in house” em Perl, para a construção de grupos de sequência teste e de construção dos HMMs. Para cada HMM construído, um GA foi escolhido, sendo que o modelo com o menor valor (portanto, mais estrito) foi escolhido para a busca dos marcadores em *Ustilaginaceae*.

Alinhamentos múltiplos foram realizados para cada marcador usando MAFFT v. 7.023 (Katoh e Standley 2013). O melhor modelo evolutivo foi selecionado de acordo com critérios do “Akaike Information Criterion” (AIC) após rodar ProtTest v.3.2 (Darriba, Taboada et al. 2011). Alinhamentos individuais foram concatenados usando FASconCAT v. 1.0 (Kück e Meusemann 2010). A filogenia foi inferida com a supermatriz usando o RAxML v. 8 (Stamatakis 2014) com partições. A sub-árvore de *Ustilaginaceae* foi extraída para análises subsequentes. Análises adicionais de inferência filogenética foram realizadas com abordagem bayesiana MrBayes (versão de desenvolvimento v3.2.7, que inclui o modelo evolutivo LG), gamma, variável ratepr, relógio uniforme, com 100.000 gerações. Árvores de espécies foram também inferidas usando FastTree (Price et al., 2009) e IQ-TREE (Nguyen et al., 2014) para a supermatriz com todos os ortólogos de cópia única, sem partições, usando o modelo WAG (Whelan and Goldman, 2001) com gamma e taxas de substituição variáveis (Yang, 1994). IQ-TREE foi usado com 1000 repetições de bootstrap.

Tabela 10 - Informações sobre os quatro marcadores de Capella-Gutierrez et al. (2014) e perfis HMM construídos para identificação nos proteomas incluídos neste trabalho.

Identificador do marcador	SGD	Descrição no Uniprot	Límiar de pontuação do HMM (GA)	Melhor modelo evolutivo
Ani0004670	YHR186C	Target of rapamycin complex 1 subunit KOG1	572,3	LG+F
Ani0004937	YMR012W	Clustered mitochondria protein 1	459,45	LG+F
Ani0005956	YJL029C	Vacuolar protein sorting-associated protein 53	300,3	JTT+F
Cal0000925	YAR007C	Replication factor A protein 1	354,4	LG+F

F - empirical base frequencies; JTT - modelo de Jones et al. (1992); LG - modelo de Le e Gascuel (2008); SGD - *Saccharomyces Genome Database* (www.yeastgenome.org/).

Evolução de tamanhos de famílias gênicas

Inicialmente, o tempo estimado de divergência para os nós na árvore filogenética de espécies foi estimado usando treePL (Smith e O'Meara, 2012). Os nós usados para calibração foram buscados usando TimeTree (Hedges et al., 2006); a idade dos nós (Milhões de Anos) correspondentes ao ancestral comum mais recente de *M. maydis* e *S. reilianum* (38-104) e de *M. maydis* e *A. flocculosa* (20-68) foram obtidos de Riess et al. (2016) e que, por sua vez, realizaram estimativas baseadas em datações por LePage et al., (1997) e Hibbett e Donoghue (1997).

O cronograma obtido foi usado no software CAFE v.4 (Hahn et al., 2005; Han et al., 2013) para análise de expansão e contração de famílias gênicas usando como entrada uma matriz com tamanhos de cada uma das famílias de CAZymes e TAPs em cada organismo da família *Ustilaginaceae*. Antes de rodar a análise no CAFE, o script caferror.py foi usado para incluir um modelo de erro associado aos rascunhos de genoma. O CAFE foi usado com o parâmetro filter, que usa parcimônia para predizer o tamanho de cada família no ancestral comum. Para cada família que apresentou ramos com mudança significativa de tamanho (*p*-value < 0,01), os *p*-values associados a cada expansão e contração foram corrigidos usando FDR no software R. Foram consideradas significativas as mudanças com *p*-value corrigido < 0,05.

Disponibilização dos dados

Os dados brutos de sequenciamento das leveduras *Moesziomyces spp.* F16C1 e F5C1 e *Pseudozyma sp.* F8B2 estão disponíveis no banco de dados SRA do NCBI associados ao BioProject PRJNA408120. Genomas foram depositados no NCBI sob identificadores NXFS00000000, NXFR00000000 e NXFQ00000000 para *Moesziomyces sp.* F16C1, *Moesziomyces sp.* F5C1 e *Pseudozyma sp.* F8B2, respectivamente.

RESULTADOS E DISCUSSÃO

O sequenciamento dos genomas de leveduras isoladas de flores de *H. psittacorum*

As leveduras basidiomicéticas isoladas de *H. psittacorum* e pertencentes ao grupo polifilético anteriormente atribuído ao gênero *Pseudozyma* foram selecionadas a partir de análise filogenética usando as regiões ITS e D1/D2 da subunidade maior do rDNA (Figura 8). Os isolados F16C1 e F5C1 foram selecionados por pertencerem a um clado irmão do que contém a levedura anteriormente atribuída à espécie *P. aphidis* e o isolado F8B2 foi selecionado por estar em um clado irmão ao da anteriormente atribuída à espécie *P. hubeiensis*. Em seguida, os seus genomas foram sequenciados neste trabalho, com a finalidade de explorar o seu potencial biotecnológico através de análises de genômica comparativa com outros membros de *Ustilaginaceae* que tiveram seus genomas sequenciados. Durante o desenvolvimento do projeto, houve reatribuição de nomes de espécies para as leveduras anteriormente classificadas como *Pseudozyma* (Wang et al., 2015) e criação de novos gêneros, como discutido na revisão bibliográfica. Desta forma, a atribuição realizada a partir de inferências filogenéticas moleculares deste trabalho é feita como apresentado na tabela 9.

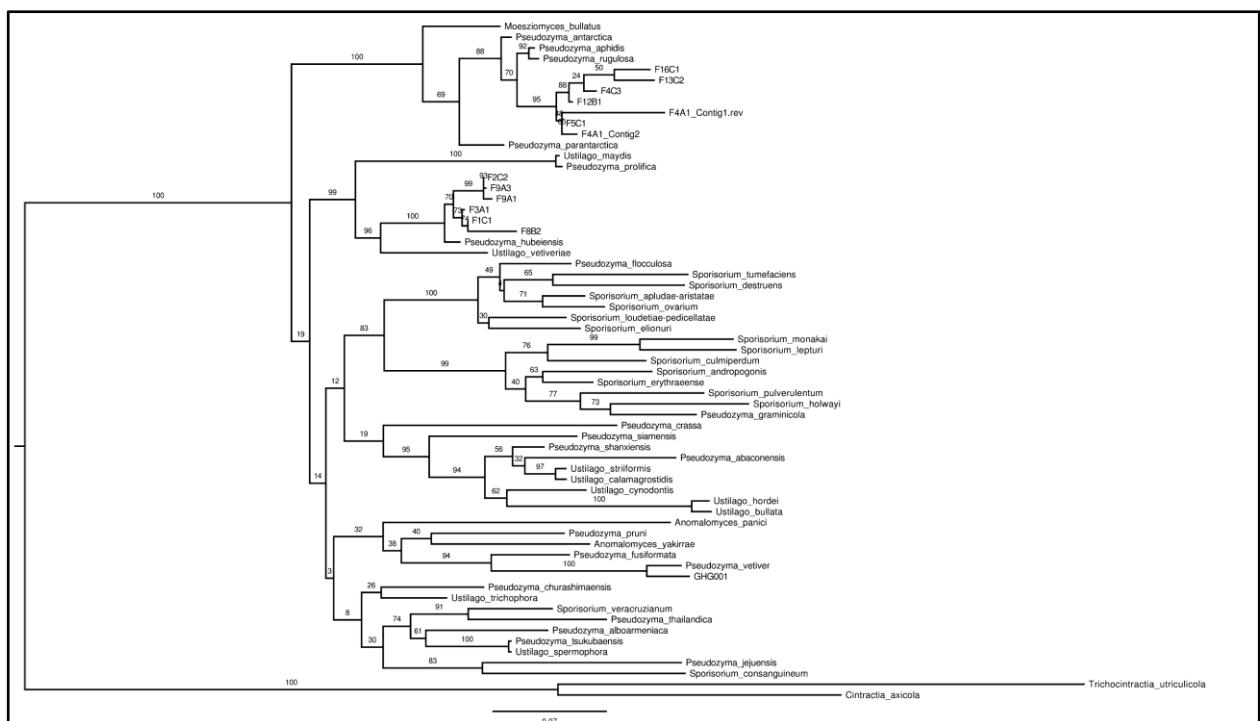


Figura 8 - Árvore filogenética Máxima Verossimilhança (ML) de isolados de *Pseudozyma* (antes do grupo polifilético ser dividido em diversos gêneros monofiléticos por Wang et al., 2015) e espécies relacionadas. Para análise filogenética, foram usadas regiões ITS1 e ITS2, 5.8S do rDNA e

subunidade D1/D2 da subunidade maior do rDNA. Os valores de bootstrap são baseados em 1000 replicatas.

Os números de “reads” antes e após a remoção das com baixa qualidade para as três leveduras estão apresentadas na tabela 11. Após comparar estatísticas de montagem (N50, tamanho da montagem e número de *scaffolds*) entre resultados de diferentes montadores, a saber Velvet, SPAdes, IDBA-UD e SAGE, o último foi selecionado. A análise de cobertura usando o script do PAGIT (Swain et al., 2012) mostrou cobertura maior de 99,8% para todos. Os genomas foram melhorados para a correção e o preenchimento de “gaps”, discontinuidades, correção de segmentos duplicados e outros inerentes da tecnologia de sequenciamento e erros do algoritmo de montagem. Além disso, etapas de “scaffolding” foram realizadas.

Tabela 11 - Número de “reads” sequenciadas e limpas para o sequenciamento do genoma de leveduras isoladas de flores de *Heliconia psittacorum*: *Moesziomyces spp.* F16C1 e F5C1 e *Pseudozyma sp.* F8B2.

Isolado	Atribuição taxonômica	Número de “reads” (pares) sequenciadas	Número de reads (pares) limpas
F16C1	<i>Moesziomyces spp.</i> F16C1	40.289.231	27.955.975 (69,38%)
F5C1	<i>Moesziomyces spp.</i> F5C1	27.799.890	19.176.348 (68,97%)
F8B2	<i>Pseudozyma sp.</i> F8B2	42.181.452	29.858.798 (70,78%)

A tabela 12 apresenta estatísticas das montagens finais de genoma para as leveduras isoladas de *Heliconia*.

Tabela 12 - Estatísticas da montagem de genomas de leveduras isoladas de flores de *H. psittacorum*: *Moesziomyces spp.* F16C1 e F5C1 e *Pseudozyma sp.* F8B2 pro tempore.

Estatística	F16C1	F5C1	F8B2
Tamanho de montagem	18.325.855	18.289.131	18.467.347
Número de scaffolds	94	118	63
N50	607.830	627.737	586.279
# Ortólogos completos	1.344 (93,46%)	1.341 (93,25%)	1.345 (93,53%)
# Ortólogos completos duplicados	80	79	88
# Ortólogos fragmentados	36	37	32
# Ortólogos ausentes	58	60	61
# Ortólogos de fungos analisados		1.438	

Com os genomas sequenciados, as regiões ITS e D1/D2 do rDNA foram extraídas e uma análise filogenética incluindo leveduras basidiomicéticas com as regiões disponíveis (Figura adicional 2) (McTaggart et al., 2012a; McTaggart et al., 2012b; Piątek et al., 2015). A análise filogenética revelou que os isolados de *Moesziomyces sp.* (F16C1 e F5C1) são proximamente relacionados com *M. aphidis*, *M. antarcticus* e *M. rugulosus*. O isolado de “*Pseudozyma*” (entre aspas, pois os membros de *Pseudozyma* serão, provavelmente em breve, reatribuídos a novos gêneros) está próximo de *Macalpinomyces arundinellae-setosae* e *P. hubeiensis*. A contagem de diferenças foi contada, portanto, entre nossos isolados e os organismos relacionados, mostrando que pode haver diferenças significativas (tabela adicional 8). O trabalho de definição destes organismos como novas espécies ou atribuição a uma espécie previamente descrita será realizado em breve, incluindo dados fisiológicos e bioquímicos das leveduras (já obtidos por nosso grupo de pesquisa).

As filogenias de espécies e a alocação das novas leveduras sequenciadas

Os fungos basidiomicetos estão divididos em três classes: Pucciniomycotina, Agaricomycotina e Ustilaginomycotina (Spatafora et al., 2017). A média do tamanho dos genomas em *Ustilaginaceae* usados neste estudo é de 18,09 Mb, maiores que os genomas de Saccharomycotina e do grupo externo, *Malassezia globosa*, e menores que genomas de Pezizomycotina (Cornell et al., 2007). Entre os basidiomicetos, a média de tamanho de genomas de *Ustilaginaceae* é menor que de Agaricomycotina e Pucciniomycotina (Riley et al., 2014); a média do número de proteínas nos proteomas da família é de 6.330, maior que de *M. globosa*.

A árvore filogenética de Basidiomicetos inclui os fungos com genoma sequenciado e usados por Capella-Gutierrez et al. (2014) (tabela adicional 5), além de membros de *Ustilaginaceae* usados nas análises apresentadas neste trabalho (Figura 9). A filogenia concorda com o consenso atual (Spatafora et al., 2017): Agaricomycotina e Ustilaginomycotina apresentam-se como grupos-irmãos, enquanto Pucciniomycotina apresenta um ancestral comum com o ancestral destas duas classes. Neste trabalho não foram incluídos membros do novo filo Entorrhizomycotina.

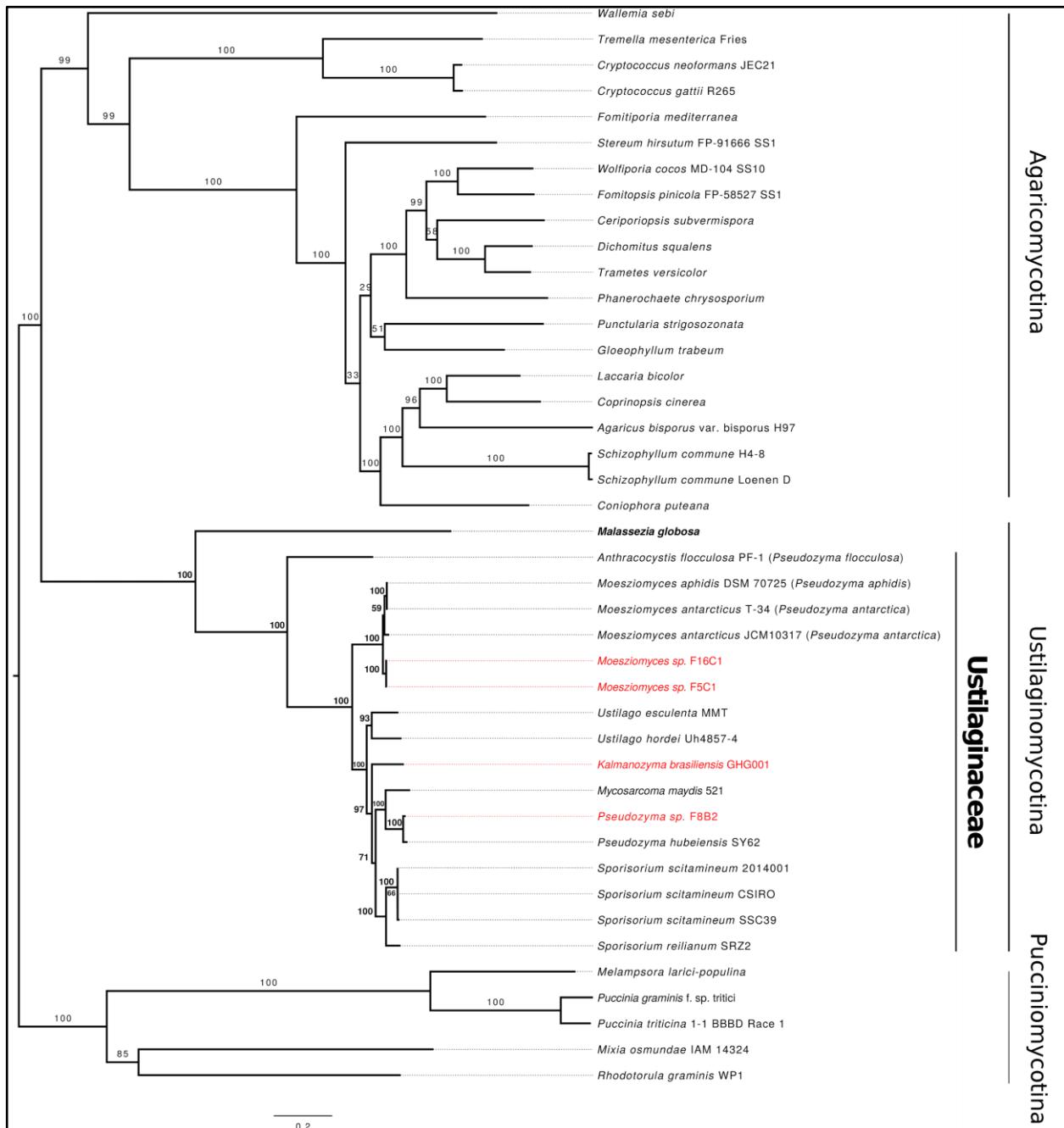


Figura 9 - Análise filogenética de Máxima Verossimilhança (ML) com supermatriz de aminoácidos e partições para os diferentes marcadores usados por Capella-Gutierrez et al., 2014 com RAxML com 1000 replicats de bootstrap de Basidiomicetos. Os modelos evolutivos para cada partição foram selecionados de acordo com os critérios da AIC (Critério de Informação Akaike) a partir dos resultados do programa ProtTest3. A família *Ustilaginaceae*, cujos membros são usados nas análises comparativas deste trabalho, está destacada em negrito. *Malassezia globosa* foi usada como grupo externo para as análises filogenética para análises de genômica comparativa, pois era o membro de Ustilaginomycotina fora de *Ustilaginaceae* com genoma sequenciado. Taxa em vermelho correspondem às leveduras com genomas montados neste trabalho.

Após a alocação dos membros de *Ustilaginaceae* entre os basidiomicetos, foram obtidas diversas filogenias incluindo os quatro marcadores (Figura 10 e Figura 11) usando diferentes métodos (raxML e a abordagem bayesiana do MrBayes)

apenas para os membros da família. A hipótese inicial foi que uma filogenia mais acurada seria obtida usando apenas membros da família, pois enquanto alguns dos marcadores não foram encontrados em alguns basidiomicetos, eles estavam presentes em todos os membros de *Ustilaginaceae* estudados e, portanto, a qualidade do alinhamento seria melhor. No entanto, como resultado foram obtidas as mesmas topologias de árvores para as árvores de *Ustilaginaceae* quando comparadas com a árvore de Basidiomicetos.

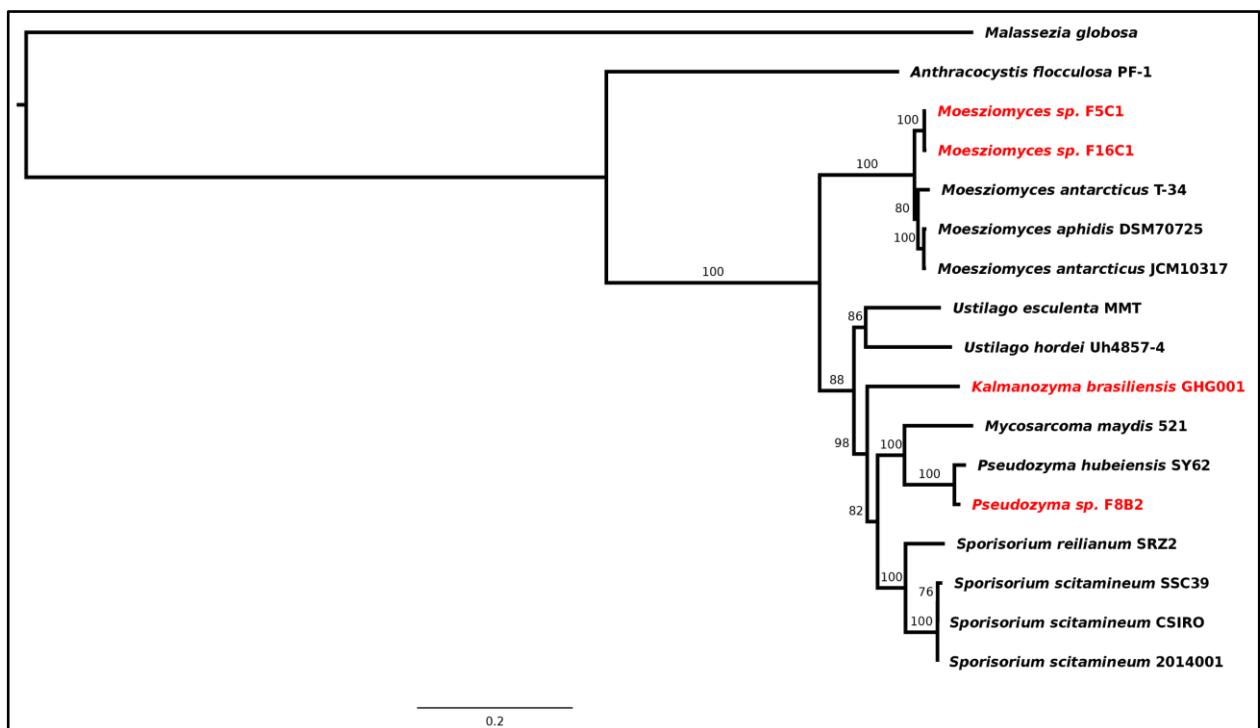


Figura 10 - Análise filogenética de Máxima Verossimilhança (ML) com supermatriz de aminoácidos e partições para os diferentes marcadores usados por Capella-Gutierrez et al., 2014 com RAxML com 1000 replicatas de bootstrap de *Ustilaginaceae*. Os modelos evolutivos para cada partição foram selecionados de acordo com os critérios da AIC (Critério de Informação Akaike) a partir dos resultados do programa ProtTest3.

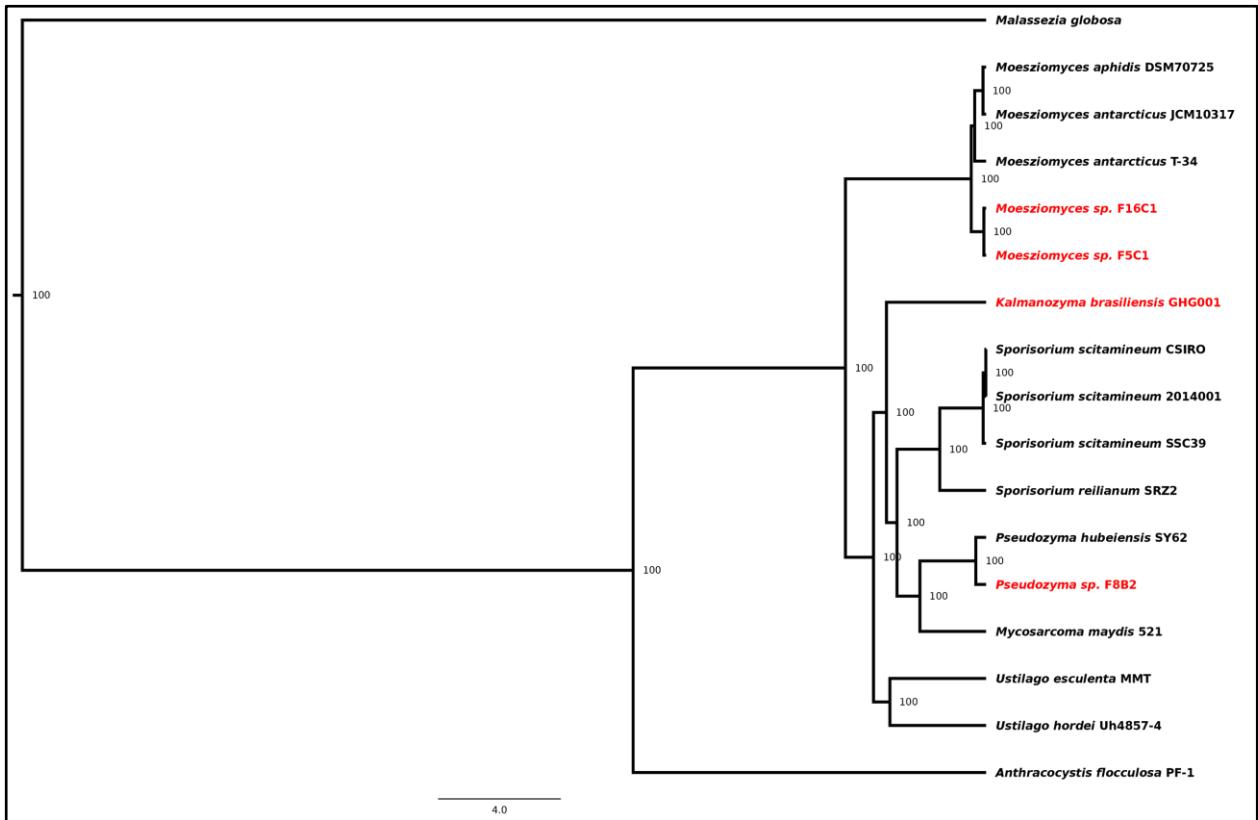


Figura 11 - MrBayes v3.2.7 usando os quatro marcadores de Capella-Gutierrez, et al. (2014). A inferência filogenética foi realizada com partições, usando os modelos evolutivos para cada marcador de acordo com a tabela 10, além de taxas de evolução gama, taxas variáveis para cada partição (ratepr=variable), e relógio uniforme. O programa foi rodado usando 100.000 gerações.

Adicionalmente, foram construídas filogenias incluindo todos os ortólogos de cópia única (Figura 12 e Figura 13) usando métodos implementados nos programas FastTree e IQ-TREE. Apesar de os quatro marcadores de Capella-Gutierrez et al. (2014) serem bons para a reconstrução de relações filogenéticas em espécies de fungos, como temos informação de todos ortólogos detectados como cópia única em *Ustilaginaceae*, reconstruímos para verificar se a mesma topologia seria recuperada.

Como resultado, foi obtida a mesma topologia para a disposição dos clados observados, exceto para *U. esculenta* e *U. hordei*. Nas filogenias com os quatro marcadores estes organismos agrupam-se em um único clado, enquanto estão agrupados em clados diferentes (com um único organismo) nas filogenias com os ortólogos de cópia única.

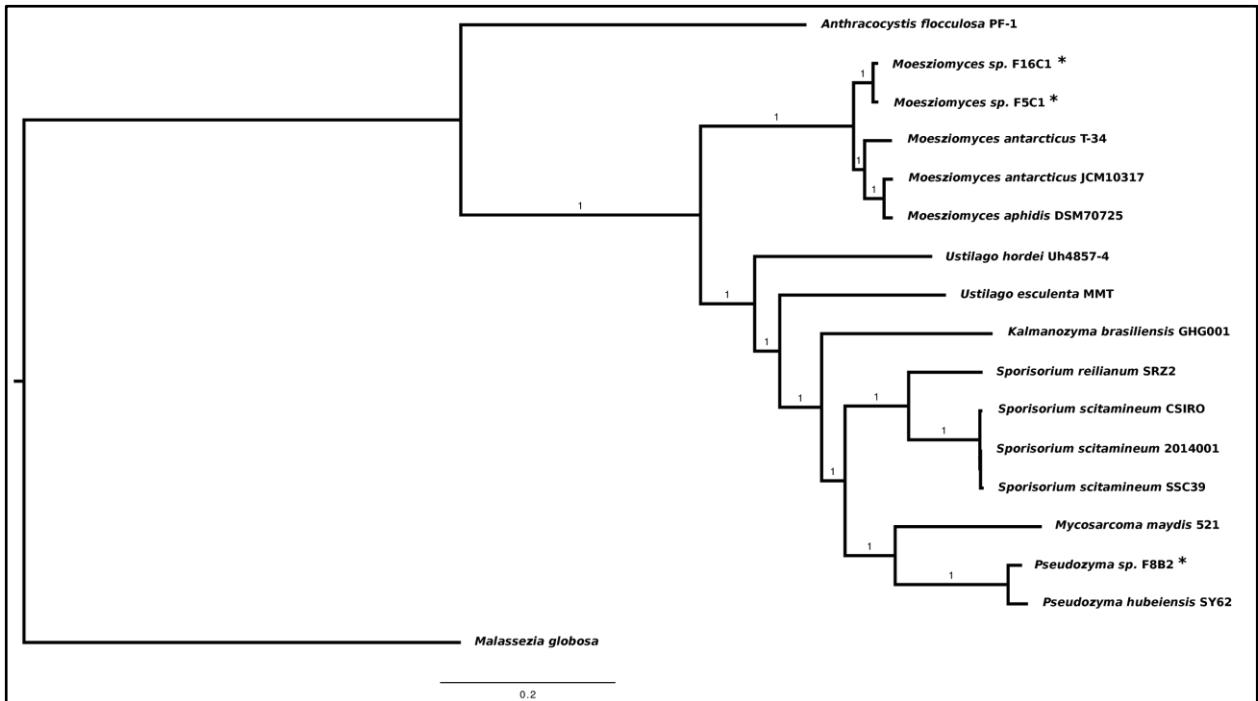


Figura 12 - Análise filogenética de Máxima Verossimilhança (ML) com supermatriz de aminoácidos usando o método implementado no FastTree e todos os ortólogos de cópia única + ortólogos de cópia única com um organismo ausente. O modelo evolutivo WAG+G foi aplicado.

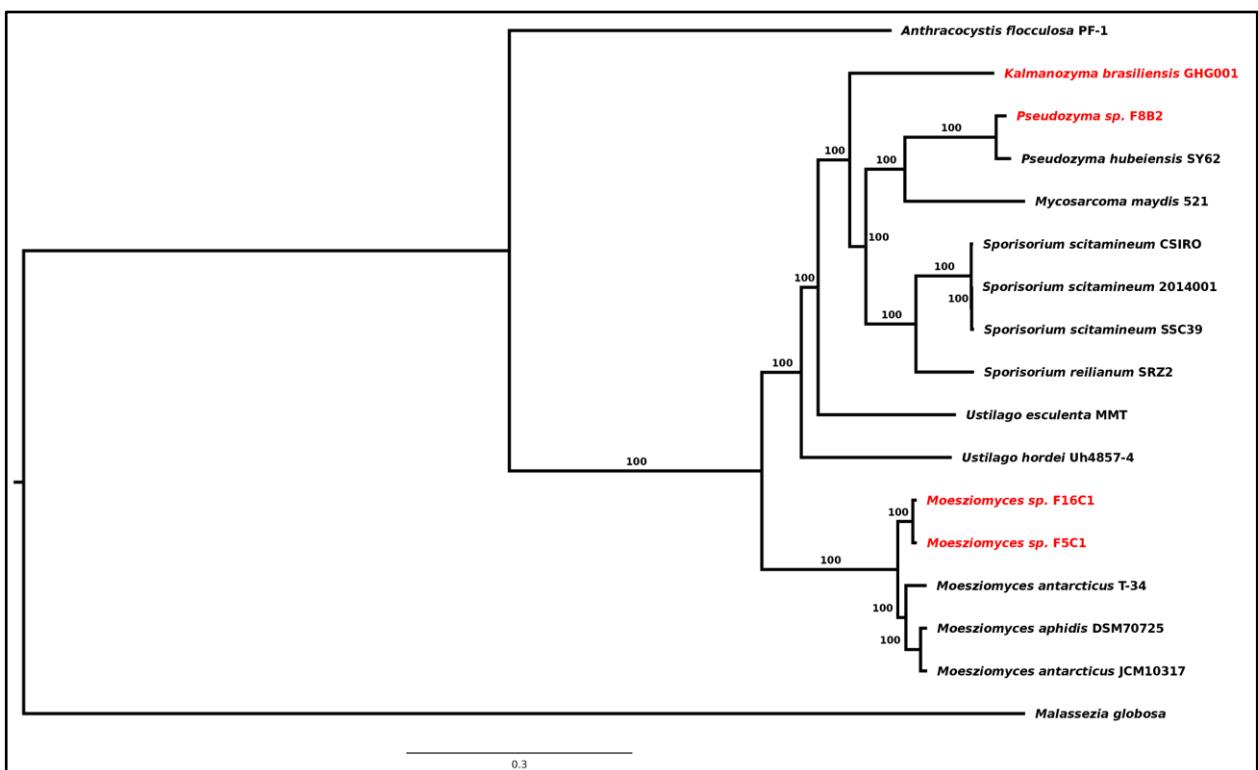


Figura 13 - Inferência usando IQ-TREE com todos os ortólogos de cópia única + ortólogos de cópia única com um organismo ausente. O modelo evolutivo WAG+G foi aplicado. Os valores de bootstrap são baseados em 1000 replicatas.

Para ganhar mais informações a partir de genomas completos, foram incluídos também os ortólogos para os quais um taxon estava ausente; nossa hipótese inicial era de que estes ortólogos ausentes seriam decorrentes principalmente da presença de genomas incompletos ou pouco explorados (“draft”). No entanto, as análises (Figura 14) mostraram que a maior parte dos genes ausentes está presente em *M. globosa*, o que era esperado por constituir o grupo externo das análises. Um dos membros da família *Moesziomyces aphidis* DSM70725 apresenta um número grande de ortólogos ausentes; para avaliar se a ausência seria decorrente da qualidade da montagem (visto que isso não acontece para outros organismos do gênero), foi feita análise com BUSCO para todos os genomas (tabela 13).

Tabela 13 - Análise com BUSCO para 1438 ortólogos de cópia única em fungos da família *Ustilaginaceae* incluídos neste trabalho.

Organismo / Genoma	BUSCO			
	Completos	Completos Duplicados	Fragmentados	Ausentes
<i>Sporisorium reilianum</i> SRZ2	1356	85	36	46
<i>Ustilago hordei</i> Uh4857-4	1309	89	46	83
<i>Mycosarcoma maydis</i> 521	1330	86	43	65
<i>Moesziomyces antarcticus</i> T-34	1326	80	44	68
<i>Pseudozyma hubeiensis</i> SY62	1350	82	37	51
<i>Anthracocystis flocculosa</i> PF-1	1331	75	54	53
<i>Kalmanozyma brasiliensis</i> GHG001	1342	85	36	60
<i>Moesziomyces aphidis</i> DSM70725	1338	76	38	62
<i>Moesziomyces antarcticus</i> JCM10317	1341	77	36	61
<i>Sporisorium</i>	1335	85	41	62

<i>scitamineum</i> 2014001				
<i>Ustilago esculenta</i> MMT	1301	78	49	88
<i>Sporisorium</i> <i>scitamineum</i> SSC39	1305	77	50	83
<i>Sporisorium</i> <i>scitamineum</i> CSIRO	1309	84	46	83

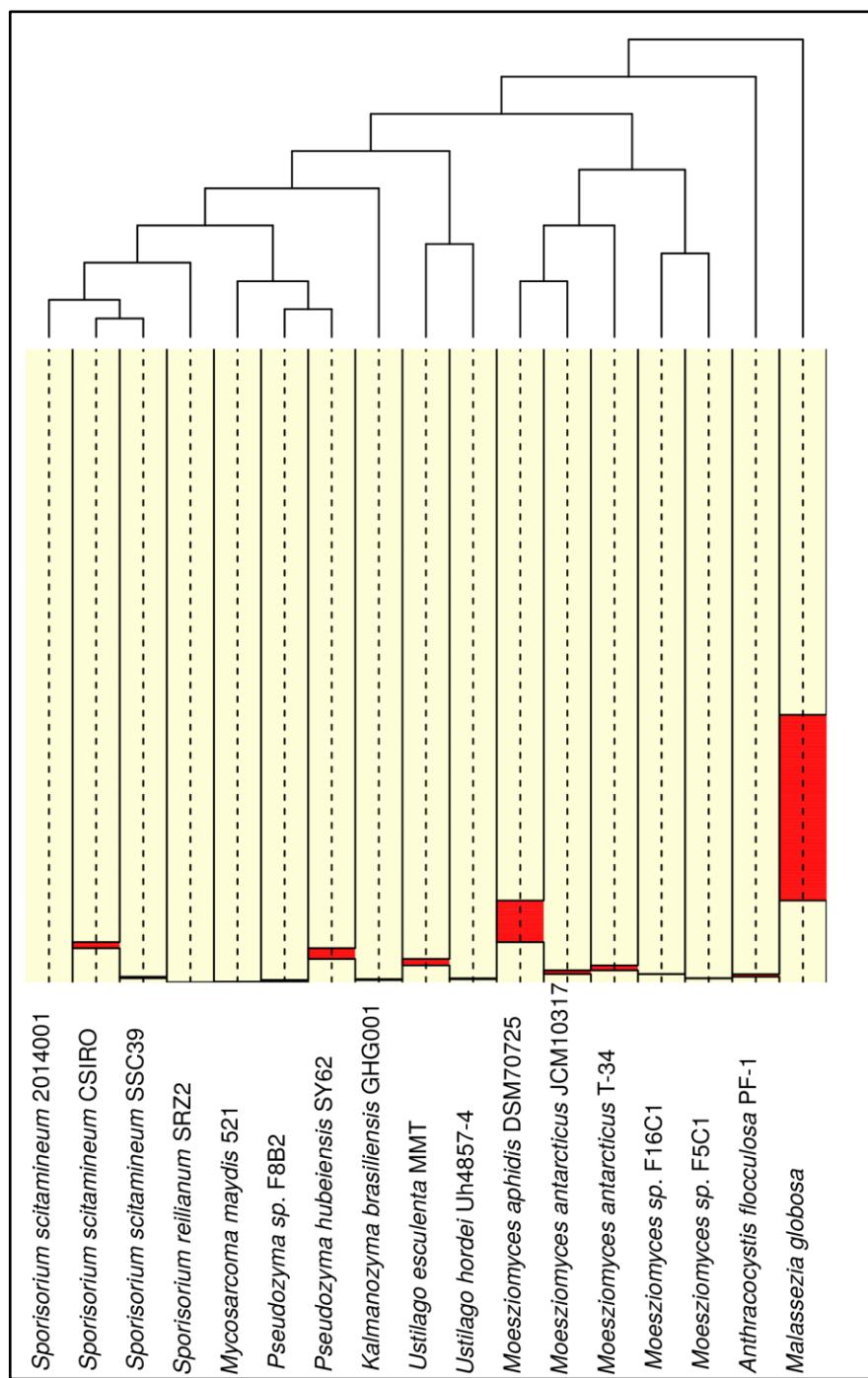


Figura 14 - Análise de presença e ausência de ortólogos de cópia única nos proteomas de *Ustilaginaceae* incluídos nas análises de genômica comparativa + *Malassezia globosa* (grupo externo).

As filogenias obtidas nas análises concordam com as realizadas previamente e incluindo leveduras basidiomicéticas amostradas da família *Ustilaginaceae*, como os trabalhos de Wang et al. (2015) e Zhao et al. (2017). Após explorar extensivamente métodos para a reconstrução da melhor filogenia de espécies de *Ustilaginaceae*, a filogenia de quatro marcadores (Figura 9) foi selecionada para as análises de contração e expansão de famílias gênicas apresentadas a seguir.

Distribuição de CAZymes e TAPs

A anotação de CAZymes e TAPs foi realizada em todos os genomas incluídos nas análises comparativas (*Ustilaginaceae* + *M. globosa*). Além disso, algumas macrovariáveis associadas aos organismos do estudo foram obtidas da literatura (tabela adicional 6), o que permite levantar hipóteses sobre a associação da presença, ausência e quantidade de CAZYmes com o estilo de vida, por exemplo.

O número de CAZYmes é aparentemente conservado em *Ustilaginaceae*, sendo bem maior que o obtido para *Malassezia globosa* (tabela 9). Além disso, a prevalência de CAZYmes em *Ustilaginaceae* é conservada (Figura 15), havendo apenas algumas famílias específicas de alguns taxa, as famílias exclusivas de *A. flocculosa*: CE2, CE12, PL3, PL4, GH106 e GH53. CE16 está presente em *A. flocculosa* e no grupo externo *M. globosa*. A família GT83 é exclusiva de *Ustilago hordei*, enquanto membros de AA9 aparecem apenas em *Sporisorium scitamineum* (ver discussão a seguir).

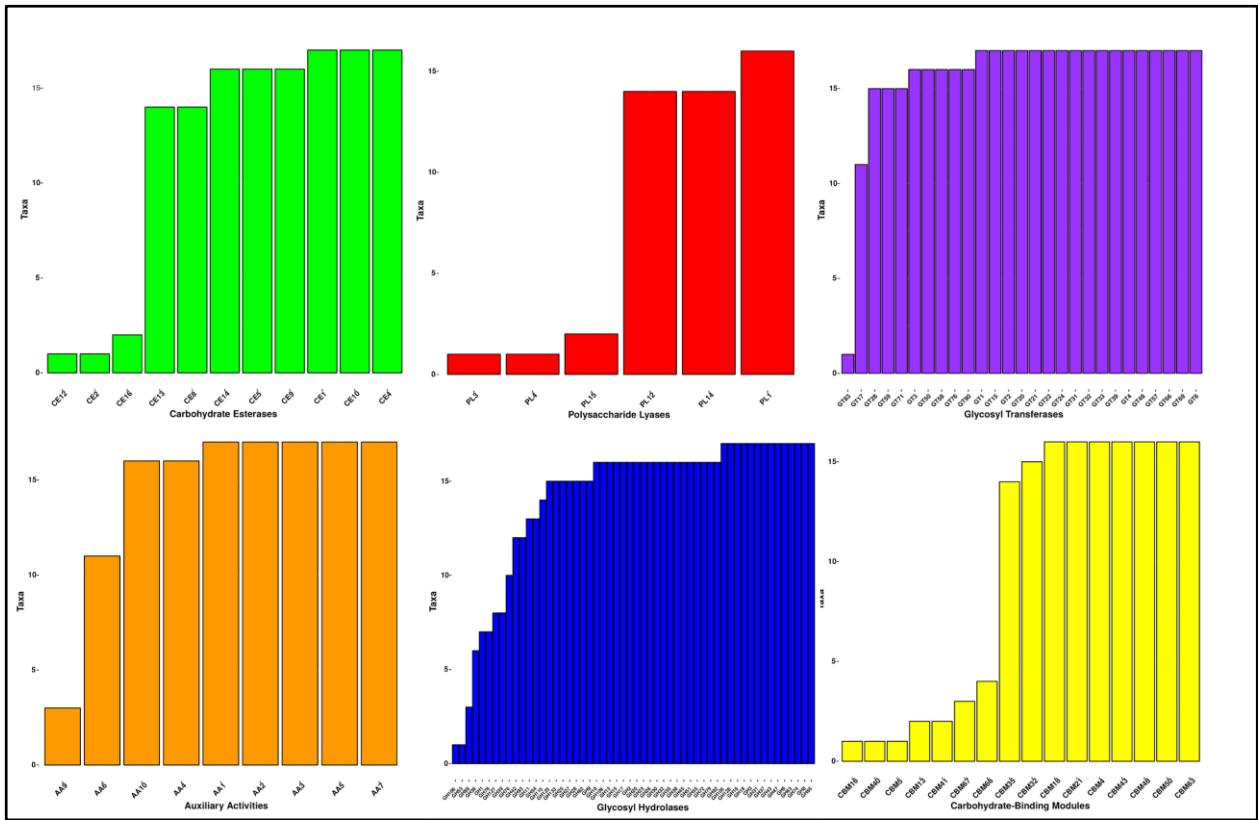


Figura 15 - Prevalência de famílias de CAZymes em *Ustilaginaceae*.

O agrupamento dos organismos de acordo com o perfil de CAZymes (Figura 16) foi realizado usando clusterização hierárquica. O agrupamento de CAZymes mostra uma diferença clara no perfil do patógeno humano *Malassezia globosa* em relação aos membros de *Ustilaginaceae*. *A. flocculosa*, embora membro da família, aparece externamente, perfil que concorda com as filogenias do grupo (figuras 10-13). *K. brasiliensis* GHG001 se agrupa com *S. reilianum* e o isolado de *Pseudozyma* sequenciado com *P. hubeiensis*, ambos agrupados com *M. maydis*, patógeno de milho. Os isolados de *Moesziomyces* sequenciados neste trabalho se agrupam com *M. antarcticus*.

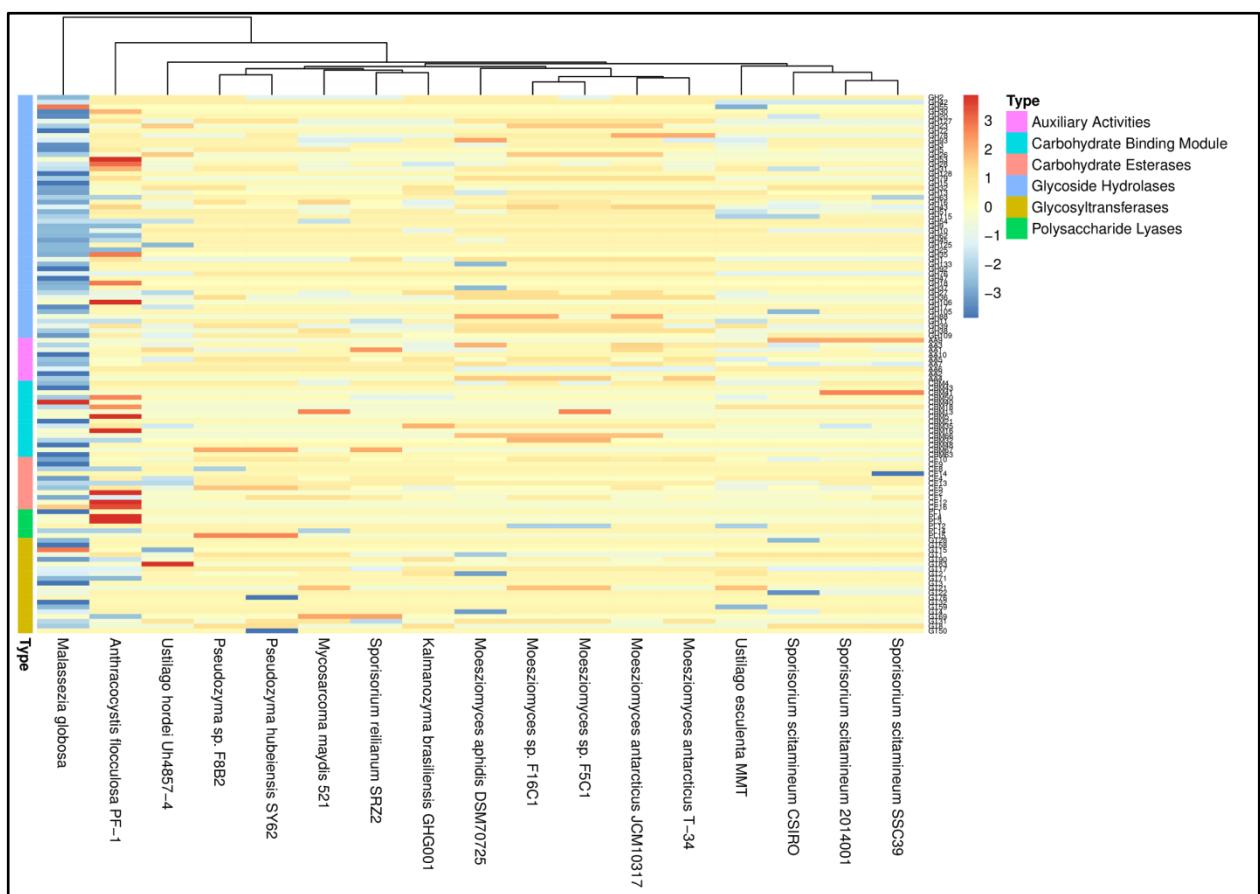


Figura 16 - Clusterização hierárquica baseada no número de proteínas atribuídas às famílias de CAZymes (ou “perfil de CAZymes”) para os organismos usados nas análises comparativas.

O perfil de TAPs (Figura 17) também foi analisado em *Ustilaginaceae*.

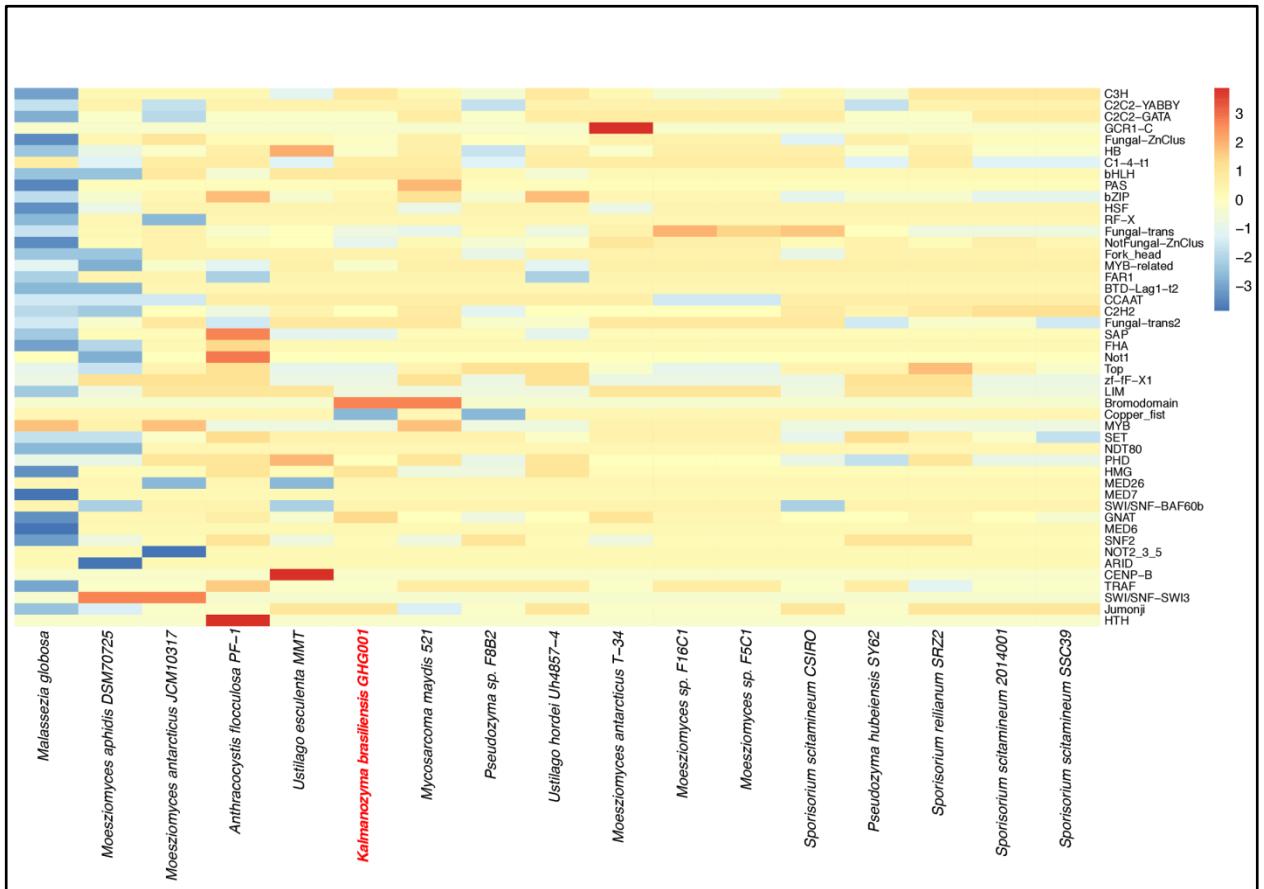


Figura 17 - Clusterização hierárquica baseada no número de proteínas atribuídas às famílias de TAPs (ou “perfil de TAPs”) para os organismos usados nas análises comparativas.

Nas buscas por LPMOs (AA 9, 10 e 11) não foram identificados membros de AA11 em nenhum dos organismos em *Ustilaginaceae*, enquanto uma cópia de gene codificando enzimas bacterianas AA10 foi encontrada em todos os organismos, exceto no grupo externo, *M. globosa*. Curiosamente, a família AA9 foi observada apenas para o clado de *Sporisorium scitamineum*, embora a família não tenha apresentado expansão significativa (seção abaixo).

Um trabalho anterior (Busk e Lange, 2015) analisou a distribuição de LPMOs em fungos, o qual incluiu apenas o genoma de *M. maydis* de *Ustilaginaceae*; os resultados do presente trabalho concordam com os de Busk e Lange, já que eles também observaram a presença apenas de AA10, entre as LPMOs. É interessante notar que o perfil de LPMOs em *Ustilaginaceae* é diferente do perfil observado em muitos outros fungos, incluindo basidiomicetos. Membros de AA9 e de AA11 são encontrados em diferentes fungos e em um número bastante variável, enquanto membros de AA10 são geralmente ausentes. O observado em *Ustilaginaceae* é o inverso.

O trabalho de Busk e Lange (2015) não inclui membros de Pucciniomycotina e o único organismo em Ustilaginomycotina é *M. maydis* - assim, seria necessários incluí-los para verificar se o perfil observado é exclusivo de *Ustilaginaceae*.

Análises de expansão de contração de famílias gênicas em *Ustilaginaceae*

As análises de contração e expansão de famílias de CAZymes e TAPs foram realizadas usando o programa CAFE. As análises levam em conta a filogenia de *Ustilaginaceae* e o processo de nascimento e morte para modelar ganho e perda de genes (Hahn et al., 2005; Han et al., 2013). Inicialmente, o programa estima o parâmetro lambda (λ), que descreve a probabilidade de ganho e perda de genes; no caso das famílias estudadas neste trabalho, o lambda estimado foi de 2,9e-5.

Os resultados apresentados mostram famílias de genes cujo p valor foi significativo, ou seja, para as quais a taxa de evolução observada foi menor ou maior que a esperada. Para cada uma destas famílias, são identificados os ramos da árvore que apresentam mudança significativa de lambda, avaliando um novo p valor (Viterbi). As famílias estão apresentadas na figura 18 (tabela adicional 7) e nós com mudança significativa de lambda incluem os organismos incluídos na análise (resultados do lado esquerda da figura) e ancestrais (nós internos, lado direito).

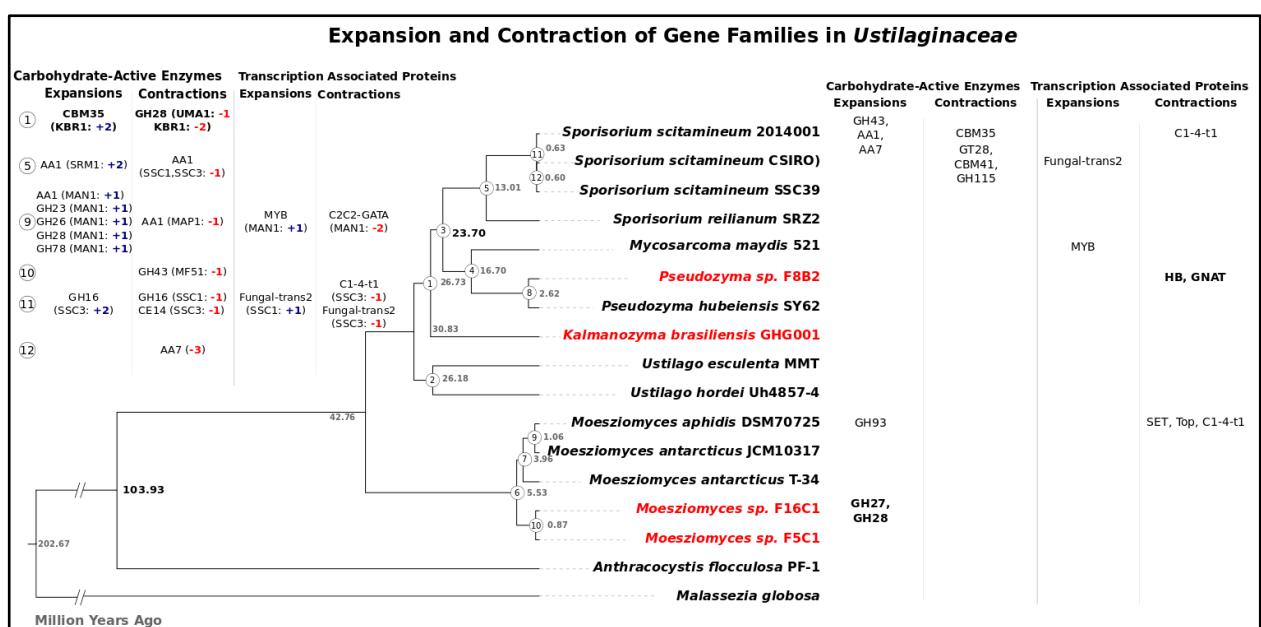


Figura 18 - Cronograma de *Ustilaginaceae* com idades de nós em Milhões de Anos e evolução de famílias de CAZymes e TAPs. A árvore filogenética de *Ustilaginaceae* com genomas sequenciados, gerada com os quatro marcadores descritos por Capella-Gutierrez et al. (2014), foi submetida a uma análise usando o programa treePL para a predição de idade dos nós na árvore. Os nós internos usados para calibração da árvore correspondem aos ancestrais comuns mais recentes a *M. maydis* e

A. flocculosa (de 38 a 104 milhões de anos) e de *M. maydis* e *S. reilianum* (de 20 a 68 milhões de anos), estimado por Riess et al. (2016) e estão destacados em negrito (calibração secundária). O cronograma foi usado para a análise de expansão e contração de famílias com o programa CAFE v.4. São apresentadas as famílias que apresentaram alteração significativa na taxa de evolução nos nós internos (esquerda) e nos membros da família (direita). Em negrito, destacam-se as famílias expandidas e contraídas em *K. brasiliensis* e nas leveduras sequenciadas neste trabalho (vide Tabela 9 para consultar os nomes correspondentes aos identificadores).

No total, 23 famílias apresentam contrações e/ou expansões significativas, entre elas 15 CAZymes e oito TAPs. Entre os organismos do estudo, apenas em *U. esculenta* e *U. hordei* (em seu MRCA) nenhuma família apresentou mudança significativa de tamanho.

Expansões de CAZymes ocorrem em um dos *S. scitamineum* (2014001), em *M. aphidis* DSM70725 e em *Moesziomyces sp.* F16C1. Mudanças significativas na taxa evolutiva de alguns em alguns nós internos (1, 5, 9 e 11) permitem identificar o aumento de tamanho de algumas famílias de CAZYmes em *S. scitamineum* SSC39, *S. reilianum*, *M. antarcticus* JCM10317 e em *K. brasiliensis* GHG001.

Curiosamente, a família de lacases e ferroxidases (AA1) apresentou o maior número de membros com alteração significativa de tamanho, com expansão e redução em diferentes membros de *Ustilaginaceae*. As funções biológicas da família são diversas, incluindo desde o papel no desenvolvimento, degradação de lignina e húmus, até a atuação durante a interação patógeno-hospedeiro (Courté et al., 2009). A variação pode indicar tanto a baixa pressão seletiva, como a especialização dos diferentes organismos.

Em *M. maydis*, apenas duas famílias apresentam mudança significativa. As famílias GH28 (galacturonases) e MYB apresentam contração e expansão, respectivamente. A contração observada é curiosa, pois enzimas da família GH28 são consideradas cruciais na degradação de pectina por patógenos de plantas, como *Sclerotinia sclerotiorum*, *Botryotinia fuckeliana* e pelo patógeno humano oportunista *Rhizopus oryzae*, que apresentam expansão de família (Zhao et al., 2013). Por outro lado, a redução de pectinases (GH28) observada em *M. maydis* pode estar associada com a menor abundância de pectina na parede celular de milho, planta por ele infectada (Mueller et al., 2008; Goulet e Saville, 2017); além disso, isso pode refletir a necessidade de manter a planta viva para formação do tumor (evitando o colapso completo do hospedeiro).

A família MYB de FTs está presente em eucariotos, incluindo fungos (Iyer et al., 2008; Todd et al., 2014), tradicionalmente estudada em plantas superiores, estando envolvida em diversos processos regulatórios (Du et al., 2015).

Curiosamente, embora o nó correspondente a *K. brasiliensis* GHG001 não tenha apresentado contração ou expansão significativa, o nó correspondente ao ancestral de *K. brasiliensis*, *Pseudozyma spp.* pro tempore, *M. maydis* e *Sporisorium spp.* apresenta mudança significativa na taxa de evolução, incluindo mudanças no tamanho de duas famílias de CAZymes em *K. brasiliensis*. A família CBM35 apresentou expansão e GH28 apresentou contração.

Em enzimas de degradação de xilano, em mananases e em exo-beta-1,3-galactanase, o CBM35 está envolvido na ligação ao xilano, galactomanano e beta-1,3-galactano, respectivamente (Bolam et al., 2004; Ichinose et al., 2005). Em *K. brasiliensis* GHG001, estes módulos estão associados a domínios de GH27 e GH43. Curiosamente, uma enzima com possível atuação em polímeros de arabinose (Peña et al., 2016), com o domínio GH27 e módulo CBM35 (evm.model.KI545858.1.137) apresenta-se diferencialmente expressa (XLORG). Por outro lado, uma GH43 (evm.model.KI545854.1.192), possivelmente xilosidase, mas sem o módulo CBM35, também apresenta-se diferencialmente expressa. Isso levanta a dúvida sobre a importância do módulo para a degradação de xilano.

A levedura parece ter perdido enzimas da família GH28. Como discutido acima, alguns patógenos apresentam expansão de GH28 em seu genoma (Zhao et al., 2013), indicando que eles podem ter capacidade de degradar pectina. Adicionalmente, genomas de basidiomicetos degradadores de biomassa comumente apresentam um grande número de genes que codificam enzimas da família GH28 (3-17 genes) (van den Brink e de Vries, 2011). No entanto, apesar de estudos terem mostrado a importância destas enzimas na infecção e virulência em patógenos de planta, nem todos os “knock-outs” de genes que codificam pectinases mostraram-se essenciais na patogenicidade (Castruita- Domínguez et al., 2014).

Caso *K. brasiliensis* GHG001 seja um organismo que mantém relações ecológicas no trato intestinal de larvas de insetos, como o nosso isolado, a ausência desta enzima pode indicar tanto que o inseto se alimenta de plantas com baixo conteúdo de pectina ou que a levedura vive em um ambiente cujos produtos de degradação de pectinas são realizados por outros organismos. Estas evidências podem reforçar também a incapacidade de infecção de plantas por *K. brasiliensis*.

No entanto, além de GH28, outras enzimas estão relacionadas com degradação de pectinas, incluindo CE8, PL1, PL2, PL3, PL9, PL10, GH78, GH88 (Zhao et al., 2013). Em *K. brasiliensis*, CE8 (metilesterase de pectina) e PL1 (liase de pectina) foram identificadas, sendo que uma cópia de PL1 está presente em todos os *Ustilaginaceae* estudados. Assim, novos estudos deverão ser realizados para melhor compreensão da degradação de pectina na levedura e possível impacto na capacidade de degradação de biomassa e no estilo de vida.

CONCLUSÕES

- As análises filogenéticas incluindo marcadores filogenômicos e informações de ortólogos de cópia única confirmam a localização dos isolados sequenciados, *Moesziomyces spp.* F16C1 e F5C1 e *Pseudozyma sp.* F8B2 pro tempore, previamente realizada usando marcadores ITS e D1/D2 (rDNA), e fornecem uma base sólida (filogenia acurada) para a realização de análises de genômica comparativa.
- As análises moleculares adicionais incluindo diversas leveduras basidiomicéticas sem genoma sequenciado, permitiram identificar as filogeneticamente próximas de *Moesziomyces spp.* F16C1 e F5C1 e *Pseudozyma sp.* F8B2 pro tempore. Análises detalhadas das diferenças em regiões ITS e da subunidade ribossomal sugerem novas espécies.
- A clusterização hierárquica das espécies baseada no perfil de CAZymes revela um agrupamento bastante similar ao esperado (grupos monofiléticos para os gêneros da família *Ustilaginaceae*). Algumas exceções de destaque incluem o agrupamento de *U. esculenta* com isolados de *S. scitamineum* e de *S. reilianum* com *K. brasiliensis*, mostrando que as diferenças na quantidade de enzimas pode ser sutis para refletir a filogenia da família de fungos.
- A prevalência de CAZymes destaca famílias únicas em organismos de *Ustilaginaceae*. Por exemplo, diversas enzimas são exclusivas de *A. flocculosa*, o que parece refletir sua divergência evolutiva em relação aos demais organismos. Outro exemplo envolve análises de LPMOs em *Ustilaginaceae*, que apresenta um perfil diferente do observado em muitos outros fungos, como a presença de AA10 e a ausência de AA9 e de AA11, mas com a presença de AA9 exclusivamente em *S. scitamineum* (três isolados).
- Análises de expansão e contração de famílias de CAZYmes e TAPs revelam possíveis adaptações de membros da família de fungos, em especial deta no fitopatógeno *M. maydis* 521 e em *K. brasiliensis* GHG001. levantam hipótese sobre a alimentação da levedura no ambiente natural na ausência de pectinases e degradação de açúcares usando enzimas GH27, possivelmente envolvida na degradação de cadeias laterais com arabinose, e GH43, possivelmente uma xilanase, podem ser mais eficientes na degradação de xilanos com a presença de módulos CBM35.

- Análises de expansão e contração nas novas leveduras sequenciadas revelam CAZymes expandidas nas famílias GH27 e GH28 (*Moesziomyces sp.* F16C1) e possíveis famílias de reguladores da transcrição contraídas, HB e GNAT (*Pseudozyma sp. pro tempore*).

CONCLUSÕES E PERSPECTIVAS

- A análise de co-regulação de enzimas e reguladores da transcrição pode levantar hipóteses sobre a regulação de CAZymes em *K. brasiliensis* GHG001 e possíveis alvos para validação.
- A exploração dos dados de RNA-seq com foco em enzimas do metabolismo de xilano e xirole se provê base para o melhoramento do modelo de redes metabólicas de *K. brasiliensis* GHG001, além de promover restrições para o modelo para simulações nas condições experimentais exploradas neste trabalho de mestrado.
- CBM35, um módulo de ligação ao carboidrato (incluindo xilano) expandido na evolução de *Ustilaginaceae*, está presente em GH43 e GH27; em particular, a GH27 com CBM35 apresenta expressão diferencial nas análises de RNA-seq de *K. brasiliensis* GHG001. Experimentos adicionais podem ser úteis na validação de CBM35 como importante na degradação de xilano.
- Apesar do destaque na literatura para outros grupos de Basidiomicetos na capacidade de degradar açúcares da parede celular de plantas, como o subfilo Agaricomycotina, os dados de ômicas apresentados neste trabalho revelam o potencial da família *Ustilaginaceae* (*Ustilaginomycotina*), como possíveis genes envolvidos na degradação de xilanos, transporte e metabolismo de xirole na levedura *K. brasiliensis* GHG001 e de possíveis especializações na degradação e regulação da expressão gênica em membros desta família de fungos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Adsul, Mukund G., Kulbhushan B. Bastawde, e Digambar V. Gokhale. "Biochemical characterization of two xylanases from yeast *Pseudozyma hubeiensis* producing only xylooligosaccharides." *Bioresource technology* 100.24 (2009): 6488-6495.
- Agger, Jane W., et al. "Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation." *Proceedings of the National Academy of Sciences* 111.17 (2014): 6287-6292.
- Altschul, Stephen F., et al. "Basic local alignment search tool." *Journal of molecular biology* 215.3 (1990): 403-410.
- Amore, Antonella, Simona Giacobbe, e Vincenza Faraco. "Regulation of cellulase and hemicellulase gene expression in fungi." *Current genomics* 14.4 (2013): 230-249.
- Augusto Corrêa dos Santos, Renato, Gustavo Henrique Goldman, e Diego Mauricio Riaño-Pachón. "ploidyNGS: visually exploring ploidy with Next Generation Sequencing data." *Bioinformatics* (2017): btx204.
- Avis, Tyler J., e Richard R. Bélanger. "Mechanisms and means of detection of biocontrol activity of *Pseudozyma* yeasts against plant-pathogenic fungi." *FEMS yeast research* 2.1 (2002): 5-8.
- Bakkeren, Guus, Jörg Kämper e Jan Schirawski. "Sex in smut fungi: structure, function and evolution of mating-type complexes." *Fungal Genetics and Biology* 45 (2008): S15-S21.
- Barbosa, Anne C., et al. "*Wickerhamiella pagnoccae* sp. nov. and *Candida tocantinsensis* sp. nov., two ascomycetous yeasts from flower bracts of *Heliconia psittacorum* (*Heliconiaceae*).". *International journal of systematic and evolutionary microbiology* 62.2 (2012): 459-464.
- Barbosa, Anne Caroline. "Diversidade e taxonomia de leveduras isoladas de flores de *Heliconia psittacorum*." (2010).
- Begerow, Dominik, Robert Bauer e Teun Boekhout. "Phylogenetic placements of ustilaginomycetous anamorphs as deduced from nuclear LSU rDNA sequences** Part 174 of the series 'Studies in Heterobasidiomycetes'." *Mycological Research* 104.1 (2000): 53-60.
- Bell, Stephen P. e Anindya Dutta. "DNA replication in eukaryotic cells." *Annual review of biochemistry* 71.1 (2002): 333-374.

- Benocci, Tiziano, et al. "Regulators of plant biomass degradation in ascomycetous fungi." *Biotechnology for Biofuels* 10.1 (2017): 152.
- Benoit, Isabelle, et al. "Closely related fungi employ diverse enzymatic strategies to degrade plant biomass." *Biotechnology for biofuels* 8.1 (2015): 107.
- Berlemont, Renaud. "Distribution and diversity of enzymes for polysaccharide degradation in fungi." *Scientific Reports* 7.1 (2017): 222.
- Bolam, David N., et al. "X4 modules represent a new family of carbohydrate-binding modules that display novel properties." *Journal of Biological Chemistry* 279.22 (2004): 22953-22963.
- Bolger, Anthony M., Marc Lohse e Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30.15 (2014): 2114-2120.
- Borges, Thuanny A., et al. "Biochemical characterization of an endoxylanase from *Pseudozyma brasiliensis* sp. nov. strain GHG001 isolated from the intestinal tract of Chrysomelidae larvae associated to sugarcane roots." *Process Biochemistry* 49.1 (2014): 77-83.
- Borges, Thuanny Andrade. "Caracterização funcional de genes importantes no metabolismo de xilose das leveduras Rhodotorula dairenensis e *Pseudozyma brasiliensis* sp. nov." (2014).
- Buitrago-Flórez, Francisco Javier, Silvia Restrepo, and Diego Mauricio Riaño-Pachón. "Identification of transcription factor genes and their correlation with the high diversity of stramenopiles." *PloS one* 9.11 (2014): e111841.
- Busk, Peter K. e Lene Lange. "Classification of fungal and bacterial lytic polysaccharide monooxygenases." *BMC genomics* 16.1 (2015): 368.
- Bussamara, Roberta, et al. "Optimal conditions for continuous immobilization of *Pseudozyma hubeiensis* (strain HB85A) lipase by adsorption in a packed-bed reactor by response surface methodology." *Enzyme research* 2012 (2012).
- Cantarel, Brandi L., et al. "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics." *Nucleic acids research* 37.suppl_1 (2008): D233-D238.
- Capella-Gutierrez, Salvador, Frank Kauff, and Toni Gabaldón. "A phylogenomics approach for selecting robust sets of phylogenetic markers." *Nucleic acids research* 42.7 (2014): e54-e54.

- Castruita-Domínguez, José P., et al. "Analysis of a polygalacturonase gene of *Ustilago maydis* and characterization of the encoded enzyme." *Journal of basic microbiology* 54.5 (2014): 340-349.
- Chikhi, Rayan, and Paul Medvedev. "Informed and automated k-mer size selection for genome assembly." *Bioinformatics* 30.1 (2013): 31-37.
- Claudel-Renard, Clotilde, et al. "Enzyme-specific profiles for genome annotation: PRIAM." *Nucleic acids research* 31.22 (2003): 6633-6639.
- Colabardini, Ana Cristina, et al. "Functional characterization of a xylose transporter in *Aspergillus nidulans*." *Biotechnology for biofuels* 7.1 (2014): 46.
- Conesa, Ana, et al. "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research." *Bioinformatics* 21.18 (2005): 3674-3676.
- Cornell, Michael J., et al. "Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi." *Genome Research* 17.12 (2007): 1809-1822.
- Courty, Pierre-Emmanuel, et al. "Phylogenetic analysis, genomic organization, and expression analysis of multi-copper oxidases in the ectomycorrhizal basidiomycete *Laccaria bicolor*." *New Phytologist* 182.3 (2009): 736-750.
- de Castro Oliveira, Juliana Velasco, et al. "Draft genome sequence of *Pseudozyma brasiliensis* sp. nov. strain GHG001, a high producer of endo-1, 4-xylanase isolated from an insect pest of sugarcane." *Genome announcements* 1.6 (2013): e00920-13.
- de Castro Oliveira, Juliana Velasco, et al. "*Pseudozyma brasiliensis* sp. nov., a xylanolytic, ustilaginomycetous yeast species isolated from an insect pest of sugarcane roots." *International journal of systematic and evolutionary microbiology* 64.6 (2014): 2159-2168.
- De Groot, Piet WJ, et al. "An endo-1, 4- β -xylanase-encoding gene from *Agaricus bisporus* is regulated by compost-specific factors." *Journal of molecular biology* 277.2 (1998): 273-284.
- Delsuc, Frédéric, Henner Brinkmann, and Philippe Hervé. "Phylogenomics and the reconstruction of the tree of life." *Nature reviews. Genetics* 6.5 (2005): 361.
- de Mendoza, Alex, et al. "Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages." *Proceedings of the National Academy of Sciences* 110.50 (2013): E4858-E4866.

- de Vries, Ronald P., et al. "Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*." *Genome biology* 18.1 (2017): 28.
- Dodd, Dylan, and Isaac KO Cann. "Enzymatic deconstruction of xylan for biofuel production." *GCB Bioenergy* 1.1 (2009): 2-17.
- Dos Reis, Mario, Philip CJ Donoghue, and Ziheng Yang. "Bayesian molecular clock dating of species divergences in the genomics era." *Nature Reviews Genetics* 17.2 (2016): 71-80.
- Du, Hai, et al. "The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion." *Scientific reports* 5 (2015).
- Duplessis, Sébastien, et al. "Obligate biotrophy features unraveled by the genomic analysis of rust fungi." *Proceedings of the National Academy of Sciences* 108.22 (2011): 9166-9171.
- Dutheil, Julien Y., et al. "A tale of genome compartmentalization: the evolution of virulence clusters in smut fungi." *Genome biology and evolution* 8.3 (2016): 681-704.
- Eastwood, Daniel C., et al. "The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi." *Science* 333.6043 (2011): 762-765.
- Ebersberger, Ingo, et al. "A consistent phylogenetic backbone for the fungi." *Molecular Biology and Evolution* 29.5 (2011): 1319-1334.
- Eddy, Sean R. "A new generation of homology search tools based on probabilistic inference." *Genome Inform.* Vol. 23. No. 1. 2009.
- Engel, Stacia R., et al. "The reference genome sequence of *Saccharomyces cerevisiae*: then and now." *G3: Genes, Genomes, Genetics* 4.3 (2014): 389-398.
- Faria, Nuno Torres, et al. "Direct xylan conversion into glycolipid biosurfactants, mannosylerythritol lipids, by *Pseudozyma antarctica* PYCC 5048 T." *Enzyme and microbial technology* 71 (2015): 58-65.
- Feldbrügge, Michael, Ronny Kellner, and Kerstin Schipper. "The biotechnological use and potential of plant pathogenic smut fungi." *Applied microbiology and biotechnology* 97.8 (2013): 3253-3265.
- Fitzpatrick, David A., et al. "A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis." *BMC evolutionary biology* 6.1 (2006): 99.

- Floudas, Dimitrios, et al. "Evolution of novel wood decay mechanisms in Agaricales revealed by the genome sequences of *Fistulina hepatica* and *Cylindrobasidium torrendii*." *Fungal Genetics and Biology* 76 (2015): 78-92.
- Floudas, Dimitrios, et al. "The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes." *Science* 336.6089 (2012): 1715-1719.
- Frank, Ari, and Pavel Pevzner. "PepNovo: de novo peptide sequencing via probabilistic network modeling." *Analytical chemistry* 77.4 (2005): 964-973.
- French, William R., et al. "Wavelet-based peak detection and a new charge inference procedure for MS/MS implemented in ProteoWizard's msConvert." *Journal of proteome research* 14.2 (2014): 1299-1307.
- Gan, Pamela, et al. "Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles." *Genome biology and evolution* 8.5 (2016): 1467-1481.
- Garamszegi, László Zsolt, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice*. Springer, 2014.
- Gazis, Romina, et al. "The genome of *Xylona heveae* provides a window into fungal endophytism." *Fungal biology* 120.1 (2016): 26-42.
- Geiser, Elena, et al. "Identification of an endo-1, 4-beta-xylanase of *Ustilago maydis*." *BMC biotechnology* 13.1 (2013): 59.
- Geiser, Elena, et al. "Prospecting the biodiversity of the fungal family *Ustilaginaceae* for the production of value-added chemicals." *Fungal Biology and Biotechnology* 1.1 (2014): 2.
- Goulet, Kristi M., and Barry J. Saville. "Carbon acquisition and metabolism changes during fungal biotrophic plant pathogenesis: insights from *Ustilago maydis*." *Canadian Journal of Plant Pathology* (2017).
- Grigoriev, Igor V., et al. "MycoCosm portal: gearing up for 1000 fungal genomes." *Nucleic Acids Research* 42.D1 (2013): D699-D704.
- Guevarra, Elvira D., and Takeshi Tabuchi. "Accumulation of itaconic, 2-hydroxyparaconic, itatartaric, and malic acids by strains of the genus *Ustilago*." *Agricultural and biological chemistry* 54.9 (1990): 2353-2358.
- Günther, Michael, et al. "The transcriptomic profile of *Pseudozyma aphidis* during production of mannosylerythritol lipids." *Applied microbiology and biotechnology* 99.3 (2015): 1375-1388.

- Gurevich, Alexey, et al. "QUAST: quality assessment tool for genome assemblies." *Bioinformatics* 29.8 (2013): 1072-1075.
- Haas, Brian J., et al. "Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments." *Genome biology* 9.1 (2008): R7.
- Hahn, Matthew W., et al. "Estimating the tempo and mode of gene family evolution from comparative genomic data." *Genome research* 15.8 (2005): 1153-1160.
- Hahn-Hägerdal, Bärbel, et al. "Towards industrial pentose-fermenting yeast strains." *Applied microbiology and biotechnology* 74.5 (2007): 937-953.
- Han, Mira V., et al. "Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3." *Molecular biology and evolution* 30.8 (2013): 1987-1997.
- Hedges, S. Blair, Joel Dudley, and Sudhir Kumar. "TimeTree: a public knowledge-base of divergence times among organisms." *Bioinformatics* 22.23 (2006): 2971-2972.
- Hibbett, D., David Grimaldi, and M. Donoghue. "Fossil mushrooms from Miocene and Cretaceous ambers and the evolution of Homobasidiomycetes." *American Journal of Botany* 84.7 (1997): 981-981.
- Himmel, Michael E., et al. "Biomass recalcitrance: engineering plants and enzymes for biofuels production." *science* 315.5813 (2007): 804-807.
- Hittinger, Chris Todd, et al. "Genomics and the making of yeast biodiversity." *Current opinion in genetics & development* 35 (2015): 100-109.
- Hoff, Katharina J., et al. "BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS." *Bioinformatics* 32.5 (2015): 767-769.
- Huang, Zhen-Bang, et al. "A novel major facilitator transporter TrSTR1 is essential for pentose utilization and involved in xylanase induction in *Trichoderma reesei*." *Biochemical and biophysical research communications* 460.3 (2015): 663-669.
- Huson, Daniel H., and Suparna Mitra. "Introduction to the analysis of environmental sequences: metagenomics with MEGAN." *Evolutionary Genomics: Statistical and Computational Methods, Volume 2* (2012): 415-429.
- Ichinose, Hitomi, et al. "An exo- β -1, 3-galactanase having a novel β -1, 3-galactan-binding module from *Phanerochaete chrysosporium*." *Journal of Biological Chemistry* 280.27 (2005): 25820-25829.

- Ilie, Lucian, et al. "SAGE: string-overlap assembly of genomes." *BMC bioinformatics* 15.1 (2014): 302.
- Iyer, Lakshminarayan M., et al. "Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes." *International journal for parasitology* 38.1 (2008): 1-31.
- Jones, David T., William R. Taylor, and Janet M. Thornton. "The rapid generation of mutation data matrices from protein sequences." *Bioinformatics* 8.3 (1992): 275-282.
- Jones, Philip, et al. "InterProScan 5: genome-scale protein function classification." *Bioinformatics* 30.9 (2014): 1236-1240.
- Kämper, Jörg, et al. "Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*." *Nature* 444.7115 (2006): 97.
- Katoh, Kazutaka, and Daron M. Standley. "MAFFT multiple sequence alignment software version 7: improvements in performance and usability." *Molecular biology and evolution* 30.4 (2013): 772-780.
- Kaupert Neto, Antonio Adalberto, et al. "Insights into the plant polysaccharide degradation potential of the xylanolytic yeast *Pseudozyma brasiliensis*." *FEMS yeast research* 16.2 (2015): foy117.
- Kawashima, Shuichi, et al. "KEGG API: A web service using SOAP/WSDL to access the KEGG system." *Genome informatics* 14 (2003): 673-674.
- Kent, W. James. "BLAT—the BLAST-like alignment tool." *Genome research* 12.4 (2002): 656-664.
- Kim, Daehwan, et al. "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." *Genome biology* 14.4 (2013): R36.
- Kobakhidze, Aza, et al. "Induction and catabolite repression of cellulase and xylanase synthesis in the selected white-rot basidiomycetes." *Annals of Agrarian Science* 14.3 (2016): 169-176.
- Konishi, Masaaki, Yuji Hatada, and Jun-ichi Horiuchi. "Draft genome sequence of the basidiomycetous yeast-like fungus *Pseudozyma hubeiensis* SY62, which produces an abundant amount of the biosurfactant mannosylerythritol lipids." *Genome announcements* 1.4 (2013): e00409-13.
- Kron, Paul, Jan Suda, and Brian C. Husband. "Applications of flow cytometry to evolutionary and population biology." *Annu. Rev. Ecol. Evol. Syst.* 38 (2007): 847-876.

- Kruse, Julia, et al. "Asexual and sexual morphs of *Moesziomyces* revisited." *IMA Fungus* (2016).
- Kuramae, Eiko E., et al. "Phylogenomics reveal a robust fungal tree of life." *FEMS Yeast Research* 6.8 (2006): 1213-1220.
- Kurtzman, Cletus P., et al. "Advances in yeast systematics and phylogeny and their use as predictors of biotechnologically important metabolic pathways." *FEMS yeast research* 15.6 (2015): foy050.
- Kuuskeri, Jaana, et al. "Time-scale dynamics of proteome and transcriptome of the white-rot fungus *Phlebia radiata*: growth on spruce wood and decay effect on lignocellulose." *Biotechnology for biofuels* 9.1 (2016): 192.
- Landell, Melissa Fontes, et al. "*Hannaella pagnoccae* sp. nov., a tremellaceous yeast species isolated from plants and soil." *International journal of systematic and evolutionary microbiology* 64.6 (2014): 1970-1977.
- Langmead, Ben, and Steven L. Salzberg. "Fast gapped-read alignment with Bowtie 2." *Nature methods* 9.4 (2012): 357-359.
- Laurie, John D., et al. "Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements." *The Plant Cell* 24.5 (2012): 1733-1745.
- Lavín, José L., et al. "Genomic analysis of two-component signal transduction proteins in basidiomycetes." *Journal of molecular microbiology and biotechnology* 18.2 (2010): 63-73.
- Le, Si Quang, and Olivier Gascuel. "An improved general amino acid replacement matrix." *Molecular biology and evolution* 25.7 (2008): 1307-1320.
- Leandro, Maria José, Paula Gonçalves, and Isabel Spencer-Martins. "Two glucose/xylose transporter genes from the yeast *Candida intermedia*: first molecular characterization of a yeast xylose-H⁺ symporter." *Biochemical Journal* 395.3 (2006): 543-549.
- Lefebvre, François, et al. "The transition from a phytopathogenic smut ancestor to an anamorphic biocontrol agent deciphered by comparative whole-genome analysis." *The Plant Cell Online* 25.6 (2013): 1946-1959.
- LePage, B., et al. "Fossil ectomycorrhizae from the Middle Eocene." *American Journal of Botany* 84.3 (1997): 410-410.

- Levasseur, Anthony, et al. "Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes." *Biotechnology for biofuels* 6.1 (2013): 41.
- Li, Heng, et al. "The sequence alignment/map format and SAMtools." *Bioinformatics* 25.16 (2009): 2078-2079.
- Li, Li, Christian J. Stoeckert, and David S. Roos. "OrthoMCL: identification of ortholog groups for eukaryotic genomes." *Genome research* 13.9 (2003): 2178-2189.
- Liao, Yang, Gordon K. Smyth, and Wei Shi. "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." *Bioinformatics* 30.7 (2013): 923-930.
- Lombard, Vincent, et al. "The carbohydrate-active enzymes database (CAZy) in 2013." *Nucleic acids research* 42.D1 (2013): D490-D495.
- Lorenz, Stefan, et al. "Genome sequence of the basidiomycetous fungus *Pseudozyma aphidis* DSM70725, an efficient producer of biosurfactant mannosylerythritol lipids." *Genome announcements* 2.1 (2014): e00053-14.
- Ludlow, Catherine L., et al. "Independent origins of yeast associated with coffee and cacao fermentation." *Current Biology* 26.7 (2016): 965-971.
- Lundell, Taina K., et al. "Genomics, lifestyles and future prospects of wood-decay and litter-decomposing basidiomycota." *Adv Bot Res* 70 (2014): 329-70.
- Marchand, G., et al. "Identification of genes potentially involved in the biocontrol activity of *Pseudozyma flocculosa*." *Phytopathology* 99.10 (2009): 1142-1149.
- Matsushika, Akinori, et al. "Characterization of non-oxidative transaldolase and transketolase enzymes in the pentose phosphate pathway with regard to xylose utilization by recombinant *Saccharomyces cerevisiae*." *Enzyme and microbial technology* 51.1 (2012): 16-25.
- McTaggart, A. R., et al. "Soral synapomorphies are significant for the systematics of the *Ustilago-Sporisorium-Macalpinomyces* complex (*Ustilaginaceae*)." *Persoonia: Molecular Phylogeny and Evolution of Fungi* 29 (2012): 63.
- McTaggart, A. R., et al. "Taxonomic revision of *Ustilago*, *Sporisorium* and *Macalpinomyces*." *Persoonia: Molecular Phylogeny and Evolution of Fungi* 29 (2012): 116.
- McTaggart, Alistair R., et al. "Mycosarcoma (*Ustilaginaceae*), a resurrected generic name for corn smut (*Ustilago maydis*) and its close relatives with hypertrophied, tubular sori." *IMA fungus* 7.2 (2016): 309-315.

- Medina, Edgar M., Gary W. Jones, and David A. Fitzpatrick. "Reconstructing the fungal tree of life using phylogenomics and a preliminary investigation of the distribution of yeast prion-like proteins in the fungal kingdom." *Journal of molecular evolution* 73.3-4 (2011): 116-133.
- Mewes, HrW, et al. "Overview of the yeast genome." *Nature* 387.6632 (1997): 7-8.
- Mhetras, Nutan, Susan Liddell, and Digambar Gokhale. "Purification and characterization of an extracellular β -xylosidase from *Pseudozyma hubeiensis* NCIM 3574 (PhXyl), an unexplored yeast." *AMB Express* 6.1 (2016): 73.
- Miller, Webb, et al. "Comparative genomics." *Annu. Rev. Genomics Hum. Genet.* 5 (2004): 15-56.
- Mimee, Benjamin, et al. "Antifungal activity of flocculosin, a novel glycolipid isolated from *Pseudozyma flocculosa*." *Antimicrobial agents and chemotherapy* 49.4 (2005): 1597-1599.
- Moon, Hee-Jung, et al. "Biotechnological production of erythritol and its applications." *Applied microbiology and biotechnology* 86.4 (2010): 1017-1025.
- Morin, Emmanuelle, et al. "Genome sequence of the button mushroom *Agaricus bisporus* reveals mechanisms governing adaptation to a humic-rich ecological niche." *Proceedings of the National Academy of Sciences* 109.43 (2012): 17501-17506.
- Morita, Tomotake, et al. "Genome sequence of the basidiomycetous yeast *Pseudozyma antarctica* T-34, a producer of the glycolipid biosurfactants mannosylerythritol lipids." *Genome announcements* 1.2 (2013): e00064-13.
- Moriya, Yuki, et al. "KAAS: an automatic genome annotation and pathway reconstruction server." *Nucleic acids research* 35.suppl 2 (2007): W182-W185.
- Moure, Andrés, et al. "Advances in the manufacture, purification and applications of xylo-oligosaccharides as food additives and nutraceuticals." *Process Biochemistry* 41.9 (2006): 1913-1923.
- Mudge, Jonathan M., and Jennifer Harrow. "The state of play in higher eukaryote gene annotation." *Nature Reviews Genetics* 17.12 (2016): 758-772.
- Mueller, Olaf, et al. "The secretome of the maize pathogen *Ustilago maydis*." *Fungal Genetics and Biology* 45 (2008): S63-S70.
- Nagy, László G., et al. "Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities." *Molecular biology and evolution* 33.4 (2015): 959-970.

- Novozhilov, Artem S., Georgy P. Karev, and Eugene V. Koonin. "Biological applications of the theory of birth-and-death processes." *Briefings in bioinformatics* (2006): 70-85.
- Ohm, Robin A., et al. "Genomics of wood-degrading fungi." *Fungal Genetics and Biology* 72 (2014): 82-90.
- Okonechnikov, Konstantin, Ana Conesa, and Fernando García-Alcalde. "Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data." *Bioinformatics* 32.2 (2015): 292-294.
- Oshlack, Alicia, Mark D. Robinson, and Matthew D. Young. "From RNA-seq reads to differential expression results." *Genome biology* 11.12 (2010): 220.
- Otto, Thomas D., et al. "Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology." *Bioinformatics* 26.14 (2010): 1704-1707.
- Parra, Genis, Keith Bradnam, and Ian Korf. "CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes." *Bioinformatics* 23.9 (2007): 1061-1067.
- Patyshakulyeva, Aleksandrina, et al. "Carbohydrate utilization and metabolism is highly differentiated in *Agaricus bisporus*." *BMC genomics* 14.1 (2013): 663.
- Paulino, Bruno N., et al. "Biotechnological production of value-added compounds by ustilaginomycetous yeasts." *Applied Microbiology and Biotechnology* (2017): 1-21.
- Paulsen, Ian T., et al. "Unified inventory of established and putative transporters encoded within the complete genome of *Saccharomyces cerevisiae*." *FEBS letters* 430.1-2 (1998): 116-125.
- Peña, María J., et al. "Structural diversity of xylans in the cell walls of monocots." *Planta* 244.3 (2016): 589-606.
- Peng, Mao, et al. "Comparative analysis of basidiomycete transcriptomes reveals a core set of expressed genes encoding plant biomass degrading enzymes." *Fungal Genetics and Biology* (2017).
- Pérez-Rodríguez, Paulino, et al. "PInTFDB: updated content and new features of the plant transcription factor database." *Nucleic acids research* 38.suppl_1 (2009): D822-D827.
- Piątek, Marcin, Matthias Lutz, and Nourou S. Yorou. "A molecular phylogenetic framework for *Anthracobystis* (Ustilaginales), including five new combinations (inter alia for the asexual *Pseudozyma flocculosa*), and description of *Anthracobystis grodzinskae* sp. nov." *Mycological Progress* 14.10 (2015): 88.

- Polizeli, M. L. T. M., et al. "Xylanases from fungi: properties and industrial applications." *Applied microbiology and biotechnology* 67.5 (2005): 577-591.
- Que, Youxiong, et al. "A global view of transcriptome dynamics during *Sporisorium scitamineum* challenge in sugarcane by RNA-Seq." *PLoS One* 9.8 (2014): e106476.
- Que, Youxiong, et al. "Genome sequencing of *Sporisorium scitamineum* provides insights into the pathogenic mechanisms of sugarcane smut." *BMC genomics* 15.1 (2014): 996.
- Riess, Kai, et al. "On the evolutionary history of *Uleiella chilensis*, a smut fungus parasite of Araucaria araucana in South America: Uleiellales ord. nov. in Ustilaginomycetes." *PloS one* 11.1 (2016): e0147107.
- Riley, Robert, et al. "Comparative genomics of biotechnologically important yeasts." *Proceedings of the National Academy of Sciences* (2016): 201603941.
- Riley, Robert, et al. "Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi." *Proceedings of the National Academy of Sciences* 111.27 (2014): 9923-9928.
- Rogowski, Artur, et al. "Evidence that GH115 α -glucuronidase activity, which is required to degrade plant biomass, is dependent on conformational flexibility." *Journal of Biological Chemistry* 289.1 (2014): 53-64.
- Rohr, Cristian O., et al. "A first insight into *Pycnoporus sanguineus* BAFC 2126 transcriptome." *PloS one* 8.12 (2013): e81033.
- Rokas, Antonis, and Chris Todd Hittinger. "Transcriptional rewiring: the proof is in the eating." *Current Biology* 17.16 (2007): R626-R628.
- Rytioja, Johanna, et al. "Plant-polysaccharide-degrading enzymes from basidiomycetes." *Microbiology and Molecular Biology Reviews* 78.4 (2014): 614-649.
- Saika, Azusa, et al. "Draft genome sequence of the yeast *Pseudozyma antarctica* type strain JCM10317, a producer of the glycolipid biosurfactants, mannosylerythritol lipids." *Genome announcements* 2.5 (2014): e00878-14.
- Scheller, Henrik Vibe, and Peter Ulvskov. "Hemicelluloses." *Annual review of plant biology* 61 (2010).
- Schirawski, Jan, et al. "Pathogenicity determinants in smut fungi revealed by genome comparison." *Science* 330.6010 (2010): 1546-1548.
- Schnittler, Martin, and Steven L. Stephenson. "Inflorescences of Neotropical herbs as a newly discovered microhabitat for myxomycetes." *Mycologia* 94.1 (2002): 6-20.

- Schoch, Conrad L., et al. "Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi." *Proceedings of the National Academy of Sciences* 109.16 (2012): 6241-6246.
- Seo, Hyuk-Seong, et al. "*Pseudozyma jejuensis* sp. nov., a novel cutinolytic ustilaginomycetous yeast species that is able to degrade plastic waste." *FEMS yeast research* 7.6 (2007): 1035-1045.
- Sharma, Krishna Kant. "Fungal genome sequencing: basic biology to biotechnology." *Critical reviews in biotechnology* 36.4 (2016): 743-759.
- Sharma, Rahul, et al. "Comparative genomics including the early-diverging smut fungus *Ceraceosorus bombacis* reveals signatures of parallel evolution within plant and animal pathogens of fungi and oomycetes." *Genome biology and evolution* 7.9 (2015): 2781-2798.
- Shelest, Ekaterina. "Transcription factors in fungi: TFome dynamics, three major families, and dual-specificity TFs." *Frontiers in genetics* 8 (2017).
- Shelest, Ekaterina. "Transcription factors in fungi." *FEMS microbiology letters* 286.2 (2008): 145-151.
- Siguier, Béatrice, et al. "First structural insights into α-L-arabinofuranosidases from the two GH62 glycoside hydrolase subfamilies." *Journal of Biological Chemistry* 289.8 (2014): 5261-5273.
- Simão, Felipe A., et al. "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs." *Bioinformatics* (2015): btv351.
- Slater, Guy St C., and Ewan Birney. "Automated generation of heuristics for biological sequence comparison." *BMC bioinformatics* 6.1 (2005): 31.
- Sloothaak, J., et al. "Aspergillus niger membrane-associated proteome analysis for the identification of glucose transporters." *Biotechnology for biofuels* 8.1 (2015): 150.
- Sloothaak, Jasper, et al. "Identification and functional characterization of novel xylose transporters from the cell factories *Aspergillus niger* and *Trichoderma reesei*." *Biotechnology for biofuels* 9.1 (2016): 148.
- Smith, Stephen A., and Brian C. O'Meara. "treePL: divergence time estimation using penalized likelihood for large phylogenies." *Bioinformatics* 28.20 (2012): 2689-2690.
- Spatafora, J. W., et al. "The Fungal Tree of Life: from Molecular Systematics to Genome-Scale Phylogenies." *Microbiology spectrum* 5.5 (2017).
- Sundar, A. Ramesh, et al. "A mini-review on smut disease of sugarcane caused by *Sporisorium scitamineum*." *Botany*. InTech, 2012.

- Swain, Martin T., et al. "A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs." *Nature protocols* 7.7 (2012): 1260.
- Taniguti, Lucas M., et al. "Complete genome sequence of *Sporisorium scitamineum* and biotrophic interaction transcriptome with sugarcane." *PloS one* 10.6 (2015): e0129318.
- Teichmann, Beate, et al. "A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*." *Molecular microbiology* 66.2 (2007): 525-533.
- Todd, Richard B., et al. "Prevalence of transcription factors in ascomycete and basidiomycete fungi." *BMC genomics* 15.1 (2014): 214.
- Tsai, Isheng J., Thomas D. Otto, and Matthew Berriman. "Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps." *Genome biology* 11.4 (2010): R41.
- van den Brink, Joost, and Ronald P. de Vries. "Fungal enzyme sets for plant polysaccharide degradation." *Applied microbiology and biotechnology* 91.6 (2011): 1477.
- Vánky, Kálmán. *Smut fungi of the world*. American Phytopathological Society, 2012.
- Vogel, John. "Unique aspects of the grass cell wall." *Current opinion in plant biology* 11.3 (2008): 301-307.
- Wang, Q-M., et al. "Multigene phylogeny and taxonomic revision of yeasts and related fungi in the Ustilaginomycotina." *Studies in mycology* 81 (2015): 55-83.
- Watanabe, Takashi, et al. "Simultaneous bioethanol distillery wastewater treatment and xylanase production by the phyllosphere yeast *Pseudozyma antarctica* GB-4 (0)." *AMB Express* 5.1 (2015): 36.
- Weierstall, Thomas, Cornelis P. Hollenberg, and Eckhard Boles. "Cloning and characterization of three genes (SUT1–3) encoding glucose transporters of the yeast *Pichia stipitis*." *Molecular microbiology* 31.3 (1999): 871-883.
- Xu, Jun, et al. "Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens." *Proceedings of the National Academy of Sciences* 104.47 (2007): 18730-18735.
- Yang, Ziheng. "Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods." *Journal of Molecular evolution* 39.3 (1994): 306-314.

- Ye, Zihong, et al. "Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome." *DNA Research* (2017).
- Yin, Yanbin, et al. "dbCAN: a web resource for automated carbohydrate-active enzyme annotation." *Nucleic acids research* 40.W1 (2012): W445-W451.
- Yoshida, Kentaro, et al. "The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine." *Elife* 2 (2013): e00731.
- Zerbino, Daniel R., and Ewan Birney. "Velvet: algorithms for de novo short read assembly using de Bruijn graphs." *Genome research* 18.5 (2008): 821-829.
- Zhao, Rui-Lin, et al. "A six-gene phylogenetic overview of Basidiomycota and allied phyla with estimated divergence times of higher taxa and a phyloproteomics perspective." *Fungal Diversity* (2017): 1-32.
- Zhao, Zhongtao, et al. "Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi." *BMC genomics* 14.1 (2013): 274.

ANEXOS

Anexo 1 - Figuras adicionais

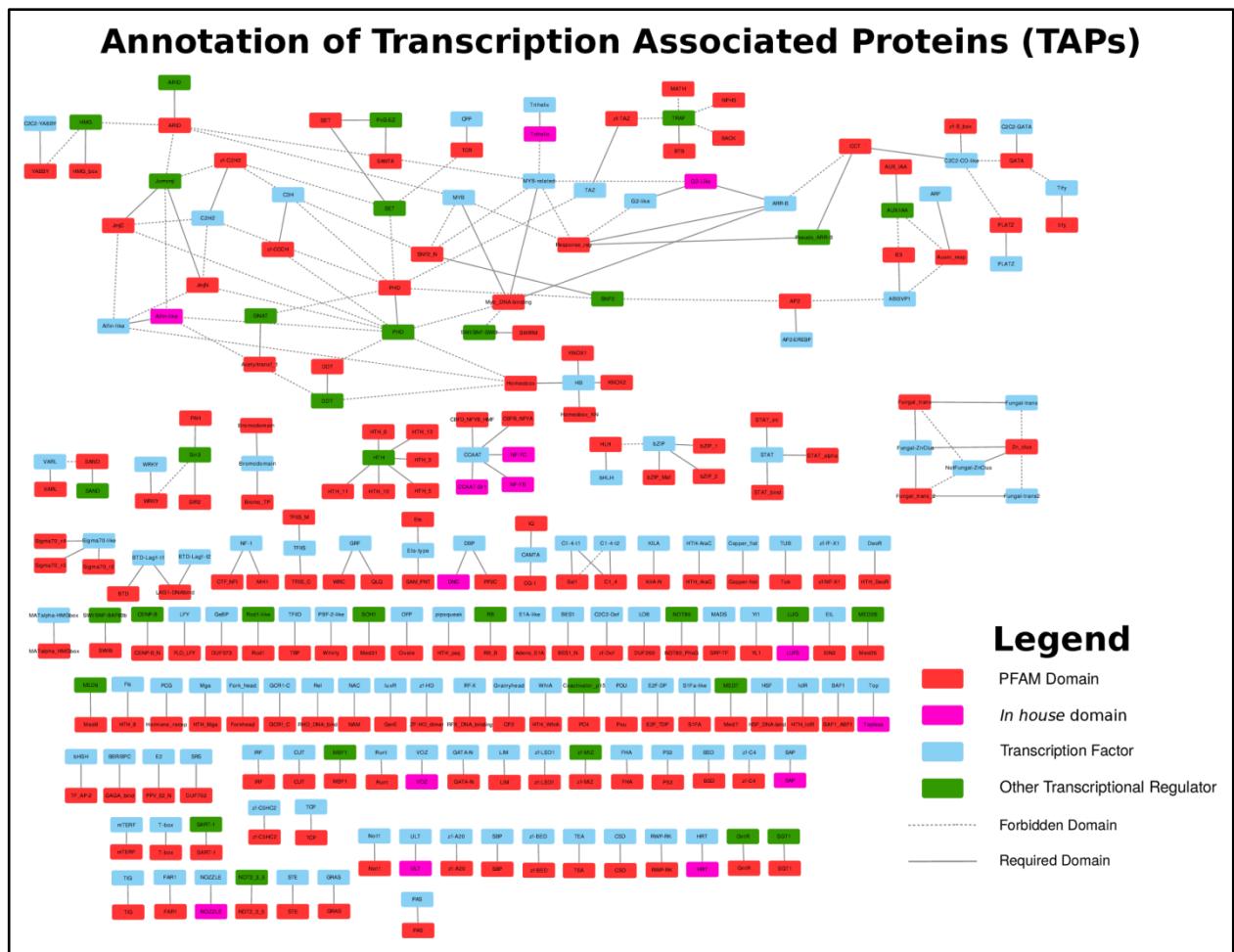


Figura adicional 1 - Diagrama representando regras para anotação de Proteínas Associadas à Transcrição (TAPs) a partir de busca de domínios conservados (“in house” e do Pfam).

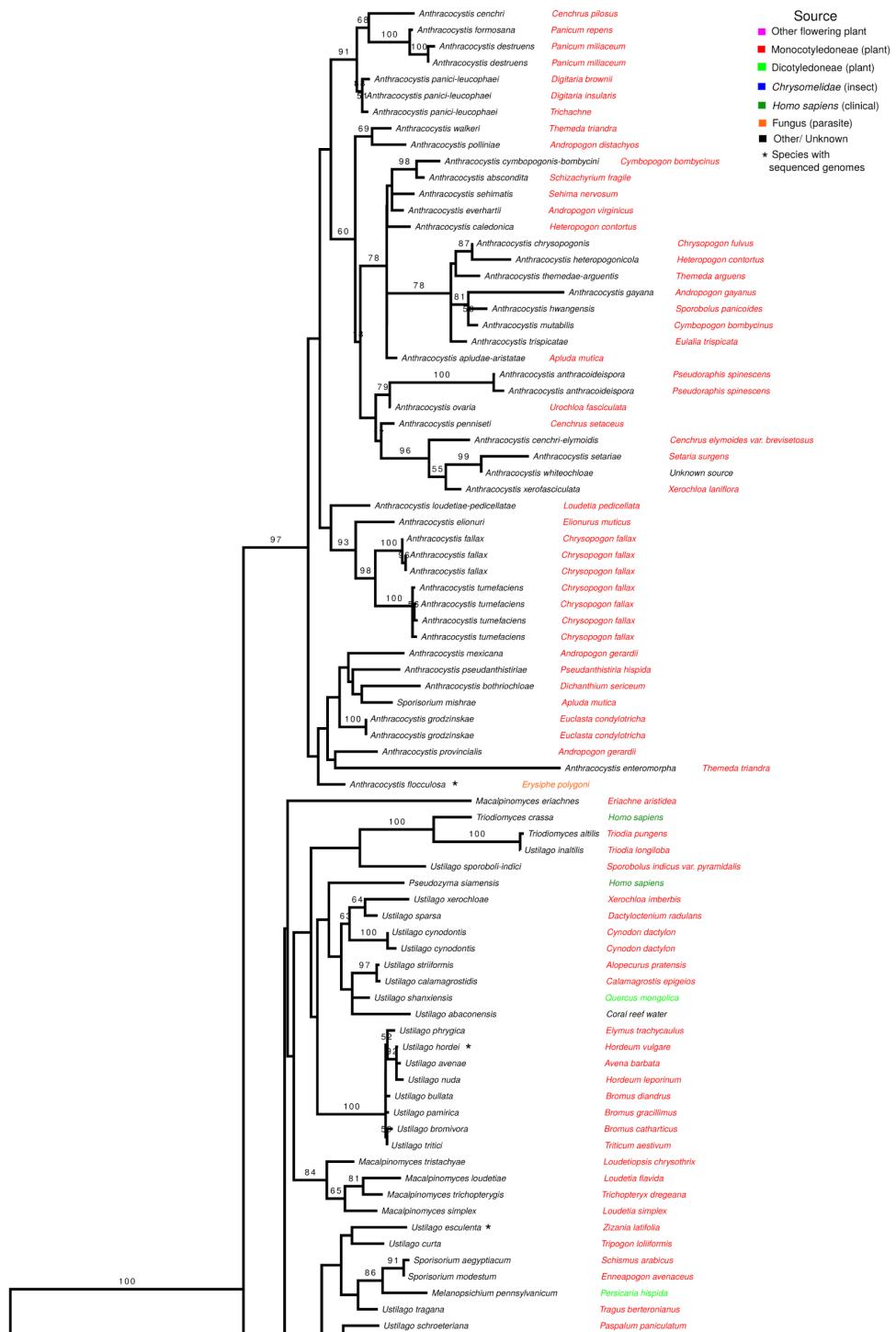




Figura adicional 2 - Filogenia de Máxima Verosimilhança (ML) com todas as espécies em Ustilaginaceae usando regiões ITS1, 5.8S, ITS2 e LSU com partições usando o programa RAxML,

levando em conta a estrutura consenso gerada pelo mxscarna e o modelo S7A, modelo GTR para frequência de nucleotídeos e distribuição GAMMA, que leva em conta taxas variáveis entre as posições. Valores de suporte são apresentados para 1000 repetições de bootstrap. Edição final com enraizamento usando “mid-point root”.

Anexo 2 - Tabelas adicionais

Tabela adicional 1 - Genes de *K. brasiliensis* GHG001 que codificam proteínas com domínios de CAZymes possivelmente envolvidas no metabolismo de xilano (Dodd e Cann, 2009; Peng et al., 2017).

Gene	Domínio	Função possível (CAZy Database)
evm.model.KI545873.1.782	GH5	endo-β-1,4-glucanase / cellulase (EC 3.2.1.4); endo-β-1,4-xylanase (EC 3.2.1.8); β-glucosidase (EC 3.2.1.21); β-mannosidase (EC 3.2.1.25); β-glucosylceramidase (EC 3.2.1.45); glucan β-1,3-glucosidase (EC 3.2.1.58); licheninase (EC 3.2.1.73); exo-β-1,4-glucanase / cellobextrinase (EC 3.2.1.74); glucan endo-1,6-β-glucosidase (EC 3.2.1.75); mannan endo-β-1,4-mannosidase (EC 3.2.1.78); cellulose β-1,4-celllobiosidase (EC 3.2.1.91); steryl β-glucosidase (EC 3.2.1.104); endoglycoceramidase (EC 3.2.1.123); chitosanase (EC 3.2.1.132); β-primeverosidase (EC 3.2.1.149); xyloglucan-specific endo-β-1,4-glucanase (EC 3.2.1.151); endo-β-1,6-galactanase (EC 3.2.1.164); hesperidin 6-O-α-L-rhamnosyl-β-glucosidase (EC 3.2.1.168); β-1,3-mannanase (EC 3.2.1.-); arabinoxylan-specific endo-β-1,4-xylanase (EC 3.2.1.-); mannan transglycosylase (EC 2.4.1.-)
evm.model.KI545873.1.249		
evm.model.KI545864.1.105		
evm.model.KI545861.1.109		
evm.model.KI545861.1.109		
evm.model.KI545863.1.116		
evm.model.KI545851.1.140		
evm.model.KI545857.1.184		
evm.model.KI545853.1.59		
evm.model.KI545851.1.601		
evm.model.KI545851.1.252	GH8	
evm.model.KI545851.1.267		
evm.model.KI545851.1.611	GH8	chitosanase (EC 3.2.1.132); cellulase (EC 3.2.1.4); licheninase (EC 3.2.1.73); endo-1,4-β-xylanase (EC 3.2.1.8); reducing-end-xylose releasing exo-oligoxyylanase (EC 3.2.1.156)
evm.model.KI545864.1.185	GH10	endo-1,4-β-xylanase (EC 3.2.1.8); endo-1,3-β-xylanase (EC 3.2.1.32); tomatinase (EC 3.2.1.-); xylan endotransglycosylase (EC 2.4.2.-)
evm.model.KI545854.1.192	GH43	β-xylosidase (EC 3.2.1.37); α-L-arabinofuranosidase

evm.model.KI545891.1.265		(EC 3.2.1.55); arabinanase (EC 3.2.1.99); xylanase (EC 3.2.1.8); galactan 1,3-β-galactosidase (EC 3.2.1.145); α-1,2-L-arabinofuranosidase (EC 3.2.1.-); exo-α-1,5-L-arabinofuranosidase (EC 3.2.1.-); [inverting] exo-α-1,5-L-arabinanase (EC 3.2.1.-); β-1,3-xylosidase (EC 3.2.1.-)
evm.model.KI545851.1.643 (com CBM4)	GH51	endoglucanase (EC 3.2.1.4); endo-β-1,4-xylanase (EC 3.2.1.8); β-xylosidase (EC 3.2.1.37); α-L-arabinofuranosidase (EC 3.2.1.55)
evm.model.KI545862.1.393 (com CBM4)		
evm.model.KI545864.1.162		
evm.model.KI545862.1.33	GH54	α-L-arabinofuranosidase (EC 3.2.1.55); β-xylosidase (EC 3.2.1.37).
evm.model.KI545864.1.182	GH62	α-L-arabinofuranosidase (EC 3.2.1.55)
evm.model.KI545851.1.465	GH3	β-glucosidase (EC 3.2.1.21); xylan 1,4-β-xylosidase (EC 3.2.1.37); β-glucosylceramidase (EC 3.2.1.45); β-N-acetylhexosaminidase (EC 3.2.1.52); α-L-arabinofuranosidase (EC 3.2.1.55); glucan 1,3-β-glucosidase (EC 3.2.1.58); glucan 1,4-β-glucosidase (EC 3.2.1.74); isopimelate-producing oligoxyloglucan hydrolase (EC 3.2.1.120); coniferin β-glucosidase (EC 3.2.1.126); exo-1,3-1,4-glucanase (EC 3.2.1.-); β-N-acetylglucosaminide phosphorylases (EC 2.4.1.-)
evm.model.KI545863.1.93		
evm.model.KI545866.1.28		
evm.model.KI545895.1.172		
evm.model.KI545851.1.142	CE1	acetyl xylan esterase (EC 3.1.1.72); cinnamoyl esterase (EC 3.1.1.-); feruloyl esterase (EC 3.1.1.73); carboxylesterase (EC 3.1.1.1); S-formylglutathione hydrolase (EC 3.1.2.12); diacylglycerol O-acyltransferase (EC 2.3.1.20); trehalose 6-O-mycolyltransferase (EC 2.3.1.122)
evm.model.KI545853.1.127		
evm.model.KI545856.1.182		
evm.model.KI545862.1.305		
evm.model.KI545862.1.415		
evm.model.KI545862.1.438		
evm.model.KI545863.1.65		
evm.model.KI545863.1.8		
evm.model.KI545864.1.22		
evm.model.KI545868.1.129		

evm.model.KI545893.1.262		
evm.model.KI545851.1.138	CE4	acetyl xylan esterase (EC 3.1.1.72); chitin deacetylase (EC 3.5.1.41); chitooligosaccharide deacetylase (EC 3.5.1.-); peptidoglycan GlcNAc deacetylase (EC 3.5.1.-); peptidoglycan N-acetylmuramic acid deacetylase (EC 3.5.1.-).
evm.model.KI545860.1.19		
evm.model.KI545862.1.347		
evm.model.KI545862.1.601		
evm.model.KI545864.1.139		
evm.model.KI545873.1.279		
evm.model.KI545873.1.615		
evm.model.KI545884.1.282		
evm.model.KI545891.1.114		
evm.model.KI545851.1.464	CE5	acetyl xylan esterase (EC 3.1.1.72); cutinase (EC 3.1.1.74)
evm.model.KI545863.1.133		
evm.model.KI545863.1.138	GH115	xylan α-1,2-glucuronidase (3.2.1.131); α-(4-O-methyl)-glucuronidase (3.2.1.-)
evm.model.KI545853.1.51	GH30	endo-β-1,4-xylanase (EC 3.2.1.8); β-glucosidase (3.2.1.21); β-glucuronidase (EC 3.2.1.31); β-xylosidase (EC 3.2.1.37); β-fucosidase (EC 3.2.1.38); glucosylceramidase (EC 3.2.1.45); β-1,6-glucanase (EC 3.2.1.75); glucuronoarabinoxylan endo-β-1,4-xylanase (EC 3.2.1.136); endo-β-1,6-galactanase (EC: 3.2.1.164); [reducing end] β-xylosidase (EC 3.2.1.-)
evm.model.KI545858.1.24		
evm.model.KI545854.1.90	GH27	α-galactosidase (EC 3.2.1.22); α-N-acetylgalactosaminidase (EC 3.2.1.49); isomaltodextranase (EC 3.2.1.94); β-L-arabinopyranosidase (EC 3.2.1.88); galactan:galactan galactosyltransferase (EC 2.4.1.-)
evm.model.KI545855.1.129		
evm.model.KI545858.1.137		
evm.model.KI545853.1.104	GH31	α-glucosidase (EC 3.2.1.20); α-galactosidase (EC 3.2.1.22); α-mannosidase (EC 3.2.1.24); α-1,3-glucosidase (EC 3.2.1.84); sucrase-isomaltase (EC 3.2.1.48) (EC 3.2.1.10); α-xylosidase (EC 3.2.1.177); α-glucan lyase (EC 4.2.2.13); isomaltosyltransferase (EC 2.4.1.-); oligosaccharide α-1,4-glucosyltransferase (EC 2.4.1.161); sulfoquinovosidase (EC 3.2.1.-)
evm.model.KI545862.1.528		
evm.model.KI545891.1.59		

evm.model.KI545864.1.53	GH74	endoglucanase (EC 3.2.1.4); oligoxyloglucan reducing end-specific cellobiohydrolase (EC 3.2.1.150); xyloglucanase (EC 3.2.1.151)
evm.model.KI545854.1.90	CBM35	Modules of approx. 130 residues. A module that is conserved in three Cellvibrio xylan-degrading enzymes binds to xylan and the interaction is calcium dependent, while a module from a Cellvibrio mannanase binds to decorated soluble mannans and mannooligosaccharides. A module in a Phanerochaete chrysosporium galactan 1,3-β-galactosidase binds to β-galactan.
evm.model.KI545858.1.137		
evm.model.KI545891.1.265		
evm.model.KI545884.1.5	AA1	Laccase / p-diphenol:oxygen oxidoreductase / ferroxidase (EC 1.10.3.2); ferroxidase (EC 1.10.3.-); Laccase-like multicopper oxidase (EC 1.10.3.-)
evm.model.KI545862.1.525	AA2	Manganese peroxidase (EC 1.11.1.13); versatile peroxidase (EC 1.11.1.16); lignin peroxidase (EC 1.11.1.14); peroxidase (EC 1.11.1.-)
evm.model.KI545868.1.86		
evm.model.KI545884.1.260		
evm.model.KI545851.1.443	AA3	Celllobiose dehydrogenase (EC 1.1.99.18); glucose 1-oxidase (EC 1.1.3.4); aryl alcohol oxidase (EC 1.1.3.7); alcohol oxidase (EC 1.1.3.13); pyranose oxidase (EC 1.1.3.10)
evm.model.KI545852.1.141		
evm.model.KI545852.1.161		
evm.model.KI545854.1.126		
evm.model.KI545862.1.274		
evm.model.KI545862.1.439		
evm.model.KI545862.1.603		
evm.model.KI545895.1.182	AA4	Vanillyl-alcohol oxidase (EC 1.1.3.38)
evm.model.KI545873.1.393		
evm.model.KI545851.1.555	AA5	Oxidase with oxygen as acceptor (EC 1.1.3.-); galactose oxidase (EC 1.1.3.9); glyoxal oxidase (EC 1.2.3.15); alcohol oxidase (EC 1.1.3.13)
evm.model.KI545873.1.272		
evm.model.KI545884.1.302		
evm.model.KI545891.1.2		

evm.model.KI545893.1.233	AA6	1,4-benzoquinone reductase (EC. 1.6.5.6)
evm.model.KI545851.1.62	AA7	Glucooligosaccharide oxidase (EC 1.1.3.-); chitooligosaccharide oxidase (EC 1.1.3.-)
evm.model.KI545853.1.164		
evm.model.KI545856.1.199		
evm.model.KI545856.1.37		
evm.model.KI545894.1.117		
evm.model.KI545894.1.245		
evm.model.KI545859.1.2	AA10	AA10 (formerly CBM33) proteins are copper-dependent lytic polysaccharide monooxygenases (LPMOs); some proteins have been shown to act on chitin, others on cellulose;

Tabela adicional 2 - Estatísticas de mapeamento de *reads* no genoma de *K. brasiliensis* GHG001 para cada experimento.

Experimento	Replicata	Número de pares de reads mapeados
Glicose	1	27,142,699 (of 27,648,043)
Glicose	2	26,191,624 (of 26,683,591)
Glicose	3	27,739,894 (of 28,296,121)
Xilose	1	24,397,465 (of 24,887,616)
Xilose	2	25,094,301 (of 25,548,492)
Xilose	3	26,648,117 (of 27,140,101)
Xilano	1	26,893,907 (of 27,402,832)
Xilano	2	28,557,090 (of 29,114,279)
Xilano	3	28,235,673 (de 28,755,635)
“Pool”	-	35.635.921 (de 36.920.105)

Tabela adicional 3 - Número de CAZymes e TAPs preditas no proteoma melhorado de *K. brasiliensis* GHG001.

Grupo	Classe	Família	Número de proteínas
TAP	FT	Bromodomain	1
TAP	FT	BSD	1
TAP	FT	C1-4-t1	1
TAP	FT	C2C2-YABBY	1
TAP	oTR	DDT	1
TAP	FT	FAR1	1

TAP	oTR	MBF1	1
TAP	oTR	MED6	1
TAP	oTR	MED7	1
TAP	FT	MYB	1
TAP	oTR	NDT80	1
TAP	FT	Not1	1
TAP	oTR	Rcd1-like	1
TAP	FT	RF-X	1
TAP	oTR	SART-1	1
TAP	oTR	SGT1	1
TAP	oTR	SOH1	1
TAP	FT	TEA	1
TAP	FT	TFIID	1
TAP	FT	TFIIS	1
TAP	FT	TIG	1
TAP	FT	YI1	1
TAP	FT	zf-C5HC2	1
TAP	oTR	zf-MIZ	1
TAP	oTR	ARID	2
TAP	FT	BTD-Lag1-t2	2
TAP	FT	Fungal-trans2	2
TAP	FT	KILA	2
TAP	FT	MADS	2
TAP	oTR	MED26	2
TAP	oTR	NOT2_3_5	2
TAP	FT	PAS	2
TAP	oTR	SWI/SNF-BAF60b	2
TAP	FT	SAP	3
TAP	oTR	TRAFF	4
TAP	FT	Fork_head	5
TAP	FT	Fungal-trans	5
TAP	FT	HSF	5
TAP	FT	LIM	5
TAP	FT	FHA	6
TAP	FT	HB	6
TAP	oTR	Jumonji	6
TAP	FT	MYB-related	6
TAP	FT	bZIP	7
TAP	oTR	PHD	7
TAP	oTR	HMG	8
TAP	FT	C2C2-GATA	9
TAP	FT	CCAAT	9
TAP	FT	C3H	10
TAP	FT	bHLH	12
TAP	oTR	SET	12
TAP	FT	Top	14
TAP	FT	C2H2	18
TAP	oTR	GNAT	20
TAP	oTR	SNF2	20
TAP	FT	Fungal-ZnClus	28

TAP	FT	NotFungal-ZnClus	29
CAZyme	AA	AA1	1
CAZyme	AA	AA10	1
CAZyme	AA	AA4	1
CAZyme	AA	AA6	1
CAZyme	CBM	CBM18	1
CAZyme	CBM	CBM21	1
CAZyme	CBM	CBM32	1
CAZyme	CBM	CBM43	1
CAZyme	CBM	CBM48	1
CAZyme	CBM	CBM63	1
CAZyme	CE	CE14	1
CAZyme	CE	CE8	1
CAZyme	CE	CE9	1
CAZyme	GH	GH10	1
CAZyme	GH	GH11	1
CAZyme	GH	GH115	1
CAZyme	GH	GH125	1
CAZyme	GH	GH133	1
CAZyme	GH	GH15	1
CAZyme	GH	GH23	1
CAZyme	GH	GH25	1
CAZyme	GH	GH26	1
CAZyme	GH	GH35	1
CAZyme	GH	GH38	1
CAZyme	GH	GH42	1
CAZyme	GH	GH54	1
CAZyme	GH	GH55	1
CAZyme	GH	GH62	1
CAZyme	GH	GH72	1
CAZyme	GH	GH76	1
CAZyme	GH	GH79	1
CAZyme	GH	GH8	1
CAZyme	GH	GH85	1
CAZyme	GH	GH9	1
CAZyme	GH	GH93	1
CAZyme	GT	GT17	1
CAZyme	GT	GT21	1
CAZyme	GT	GT24	1
CAZyme	GT	GT28	1
CAZyme	GT	GT3	1
CAZyme	GT	GT33	1
CAZyme	GT	GT48	1
CAZyme	GT	GT50	1
CAZyme	GT	GT58	1
CAZyme	GT	GT59	1
CAZyme	GT	GT66	1
CAZyme	GT	GT76	1
CAZyme	PL	PL1	1
CAZyme	PL	PL12	1

CAZyme	PL	PL14	1
CAZyme	CBM	CBM4	2
CAZyme	CBM	CBM50	1
CAZyme	CE	CE13	2
CAZyme	CE	CE5	2
CAZyme	GH	GH105	2
CAZyme	GH	GH17	2
CAZyme	GH	GH2	2
CAZyme	GH	GH20	2
CAZyme	GH	GH30	2
CAZyme	GH	GH37	2
CAZyme	GH	GH43	2
CAZyme	GH	GH63	2
CAZyme	GH	GH74	1
CAZyme	GT	GT1	2
CAZyme	GT	GT15	2
CAZyme	GT	GT31	2
CAZyme	GT	GT32	2
CAZyme	GT	GT69	2
CAZyme	GT	GT71	2
CAZyme	CBM	CBM35	3
CAZyme	GH	GH27	3
CAZyme	GH	GH31	3
CAZyme	GH	GH32	3
CAZyme	GH	GH45	3
CAZyme	GH	GH47	3
CAZyme	GH	GH51	3
CAZyme	GH	GH92	3
CAZyme	GT	GT39	3
CAZyme	GT	GT8	3
CAZyme	AA	AA2	3
CAZyme	AA	AA5	4
CAZyme	GH	GH18	4
CAZyme	GH	GH3	4
CAZyme	GT	GT20	3
CAZyme	GT	GT22	4
CAZyme	GT	GT4	4
CAZyme	GT	GT57	2
CAZyme	GH	GH109	5
CAZyme	GH	GH128	5
CAZyme	GH	GH13	5
CAZyme	AA	AA7	6
CAZyme	GT	GT90	6
CAZyme	AA	AA3	8
CAZyme	CE	CE4	9
CAZyme	CE	CE1	11
CAZyme	GT	GT2	11
CAZyme	GH	GH5	11
CAZyme	GH	GH16	13
CAZyme	CE	CE10	28

Tabela adicional 4 - Anotação de genes de *K. brasiliensis* associados a mapas do KEGG mencionados neste trabalho. A anotação de genes pertencentes a Sistemas de Dois Componentes e Replicação de DNA foi complementada com dados da literatura (Lavín et al., 2010; Bell e Dutta, 2002) e de anotação (transferência por ortologia) de *M. maydis*.

Sistema de dois componentes (TCS)						
	Gene em <i>M. maydis</i>	Gene em <i>K. brasiliensis</i> GHG001	Mapa do KEGG	Anotação funcional	DEG	Grupo
Hybrid HK	UM02739.1 KIS69408.1 (Tco1/Nik1)	evm.model.KI545891.1.60	map02020	Bacterial sensor protein C-terminal signature cheY-homologous receiver domain HAMP domain HAMP domain profile. HAMP (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases) domain His Kinase A (phospho-acceptor) domain His Kinase A (phosphoacceptor) domain Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase Histidine kinase domain profile. Histidine kinase-like ATPases Response regulator receiver domain Response regulatory domain profile.	Não	-
	UM05732.1 KIS66951.1 (Tco3/Phy1)	evm.model.KI545860.1.72	-	Bacterial sensor protein C-terminal signature cheY-homologous receiver domain Domain present in phytochromes and cGMP-specific phosphodiesterases. GAF domain His Kinase A (phospho-acceptor) domain His Kinase A (phosphoacceptor) domain Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase Histidine kinase domain profile. Histidine kinase-like ATPases PAS fold Phytochrome chromophore attachment site domain profile.	Não	-

				Phytochrome region Response regulator receiver domain Response regulatory domain profile.		
	UM04773.1 KIS6 6709.1 (Tco6)	evm.model.KI545857.1.7 5	-	Bacterial sensor protein C-terminal signature cheY-homologous receiver domain His Kinase A (phospho-acceptor) domain His Kinase A (phosphoacceptor) domain Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase Histidine kinase domain profile. Histidine kinase-like ATPases Response regulator receiver domain Response regulatory domain profile.	Não	-
	UM04907.1 KIS6 7034.1 (Tco7)	evm.model.KI545854.1.1 78	-	Bacterial sensor protein C-terminal signature cheY-homologous receiver domain Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase Histidine kinase domain profile. Histidine kinase-like ATPases Response regulator receiver domain Response regulatory domain profile.	Não	-
	UM03269.1 KIS6 8698.1	evm.model.KI545893.1.1 22	-	Bacterial sensor protein C-terminal signature cheY-homologous receiver domain His Kinase A (phospho-acceptor) domain His Kinase A (phosphoacceptor) domain Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase Histidine kinase domain profile. Histidine kinase-like ATPases Motif C-terminal to PAS motifs (likely to contribute to	Não	-

				PAS structural domain) PAC domain profile. PAS domain PAS fold PAS repeat profile. Response regulator receiver domain Response regulatory domain profile. sensory_box: PAS domain S-box protein		
HPt protein	UM05661.1 KIS6 6875.1 (Ypd1)	evm.model.KI545860.1.8 1	map02020	Histidine-containing phosphotransfer (HPt) domain profile; Hpt domain	Não	-
RR	UM00239.1 KIS7 1807.1 (Ssk1)	evm.model.KI545851.1.2 55	map02020	cheY-homologous receiver domain Response regulator receiver domain Response regulatory domain profile	Não	-
	UM03346.1 KIS6 8780.1 (Skn7)	evm.model.KI545893.1.4 2	map02020	cheY-homologous receiver domain heat shock factor Heat shock factor (HSF) domain signature HSF-type DNA-binding Response regulator receiver domain Response regulatory domain profile.	Não	-
	UM01492.1 KIS7 0322.1 (Rim15)	evm.model.KI545862.1.5 5	-	AGC-kinase C-terminal domain profile. cheY-homologous receiver domain Extension to Ser/Thr-type protein kinases Protein kinase domain Protein kinase domain profile. Response regulator receiver domain Response regulatory domain profile. Serine/Threonine protein kinases active-site signature. Serine/Threonine protein kinases, catalytic domain	Não	-
	UM04506.1 KIS6 7406.1	-	-		-	-
	UM05594.1 KIS6 6605.1	evm.model.KI545859.1.1 52	-	cheY-homologous receiver domain Response regulator receiver	Sim	XLARG

				domain Response regulatory domain profile.		
	KIS68438.1	evm.model.KI545852.1.12	map02020	Glutamine synthetase, beta-Grasp domain Glutamine synthetase, catalytic domain Glutamine synthetase signature 1.	Sim	XLORG
	KIS68485.1	evm.model.KI545852.1.205	map02020	AcCoA-C-Actrans: acetyl-CoA C-acetyltransferase Thiolase, C-terminal domain Thiolase, N-terminal domain Thiolases acyl-enzyme intermediate signature. Thiolases signature 2.	Não	-
	KIS67612.1	evm.model.KI545855.1.199	map02020	Alkaline phosphatase Alkaline phosphatase active site. Alkaline phosphatase homologues Alkaline phosphatase signature C-terminal region of a signal peptide. Hydrophobic region of a signal peptide. N-terminal region of a signal peptide. Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region. Signal peptide region	Não	-
	KIS67532.1	evm.model.KI545856.1.137	map02020	Cytochrome C1 family Cytochrome C1 signature Cytochrome c family profile. Region of a membrane-bound protein predicted to be embedded in the membrane. Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm. Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	Sim	XLORG
	KIS68243.1	evm.model.KI545861.1.104	map00190 ,map02020	Rieske [2Fe-2S] domain Rieske [2Fe-2S] iron-sulfur domain profile. Rieske 2Fe-2S subunit signature Rieske_proteo: ubiquinol-	Sim	XLORG

				cytochrome c reductase, iron-sulfur subunit Ubiquinol cytochrome reductase transmembrane region		
	KIS68276.1	evm.model.KI545861.1.137	map02020	PhoD-like phosphatase Region of a membrane-bound protein predicted to be embedded in the membrane. Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm. Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	Não	-
	KIS70374.1	evm.model.KI545862.1.109	map02020	Protein kinase domain Protein kinase domain profile. Protein kinases ATP-binding region signature. Serine/Threonine protein kinases active-site signature. Serine/Threonine protein kinases, catalytic domain	Não	-
	KIS70696.1	evm.model.KI545862.1.1418	map02020	C-terminal region of a signal peptide. Hydrophobic region of a signal peptide. N-terminal region of a signal peptide. PhoD-like phosphatase Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region. Signal peptide region	Não	-
	KIS65691.1	evm.model.KI545866.1.162	map00190, map02020	Cytochrome oxidase assembly protein Region of a membrane-bound protein predicted to be embedded in the membrane. Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm. Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	Não	-

	KIS69373.1	evm.model.KI545891.1.93	map02020	Class IA and IB cytochrome C signature Cytochrome c Cytochrome c family profile.	Sim	XLORG
	KIS66234.1	evm.model.KI545894.1.250	map02020	Glutamine synthetase, catalytic domain	Não	-

Replicação de DNA

	Element	Gene em <i>K. brasiliensis</i>	Mapa do KEGG	Anotação funcional	DEG	Grupo
Pre-replicative complex (RC)	ORC (Origin recognition complex)	evm.model.KI545855.1.53	-	Origin recognition complex subunit 6 (ORC6)	Não	-
	Cdc6p	-	-	-	-	-
	Cdt1p	evm.model.KI545894.1.96	Map00230, map00730, map01100	Helicase conserved C-terminal domain; Helicase, C-terminal; DNA replication factor CDT1 like; CDT1 Geminin-binding domain-like; DEAD/DEAH box helicase; DNA/RNA helicase, DEAD/DEAH box type, N-terminal; Domain of unknown function (DUF1998)	Não	-
	Mcm2-7p	evm.model.KI545866.1.52	map03030	MCM N-terminal domain; MCM2/3/5 family	Sim	XLAIG
	Mcm2-7p	evm.model.KI545851.1.382	map03030	MCM2/3/5 family; Mini-chromosome maintenance, DNA-dependent ATPase; MCM N-terminal domain	Não	-
	Mcm2-7p	evm.model.KI545854.1.3	map03030	MCM2/3/5 family; Mini-chromosome maintenance, DNA-dependent ATPase; MCM N-terminal domain	Não	-
	Mcm2-7p	evm.model.KI545862.1.1	map03030	Mini-chromosome maintenance protein 2; DNA replication licensing factor Mcm2; MCM N-terminal domain; MCM N-terminal domain; MCM2/3/5 family	Não	-
	Mcm2-7p	evm.model.KI545862.1.205	map03030	MCM N-terminal domain; MCM N-terminal domain; MCM2/3/5 family	Não	-
	Mcm2-7p	evm.model.KI545862.1.239	map03030	MCM2/3/5 family; Mini-chromosome maintenance,	Não	-

				DNA-dependent ATPase; MCM N-terminal domain		
	Mcm2-7p	evm.model.KI545884.1.257	-	Primase zinc finger; Zinc finger, Mcm10/DnaG-type	Não	-
	CRM1	evm.model.KI545892.1.40	-	CRM1 C terminal; CRM1 C-terminal domain; Exportin 1-like protein; Exportin-1/Importin-beta-like; Importin-beta N-terminal domain; Importin-beta, N-terminal domain	Não	-
	CDK	evm.model.KI545893.1.178	-	Cyclin-dependent kinase regulatory subunit	Não	-
Transition to replication	Mcm10	evm.model.KI545884.1.257	-	Primase zinc finger	Não	-
	Cdc45	evm.model.KI545860.1.10		CDC45-like protein	Não	-
	Sld3	evm.model.KI545853.1.103	-	DNA replication regulator SLD3	Não	-
	Replication protein A	evm.model.KI545891.1.232	map03030	Replication protein A C-terminal; Replication protein A, C-terminal; OB-fold nucleic acid binding domain; OB-fold nucleic acid binding domain, AA-tRNA synthetase-type	Não	-
	Replication factor C; ATPase family associated with various cellular activities (AAA)	evm.model.KI545851.1.233	map03030	ATPase family associated with various cellular activities (AAA); Replication factor C C-terminal domain	Não	-
	DNA primase, small subunit	evm.model.KI545893.1.54	map03030	Eukaryotic and archaeal DNA primase small subunit	Sim	XLAIG
	Replication factor C; DNA polymerase III, clamp loader complex, gamma/delt	evm.model.KI545851.1.565	map03030	Replication factor C C-terminal domain; Replication factor C, C-terminal domain; DNA polymerase III, delta subunit	Não	-

	a/delta subunit					
	ATPase family associated with various cellular activities (AAA)	evm.model.KI545852.1.22	map03030	Replication factor RFC1 C terminal domain; DNA replication factor RFC1, C-terminal; BRCA1 C Terminus (BRCT) domain; BRCT domain; ATPase family associated with various cellular activities (AAA)	Não	-
	DNA polymerase alpha/epsilon subunit B	evm.model.KI545854.1.107	map03030	DNA polymerase alpha/epsilon subunit B	Não	-
	DNA replication factor Dna2	evm.model.KI545854.1.43	map03030	AAA domain; DNA replication factor Dna2	Não	-
	Replication factor A	evm.model.KI545855.1.150	map03030	Replication factor A protein	Não	-
	Replication factor C	evm.model.KI545856.1.128	map03030	Replication factor C C-terminal domain; Replication factor C, C-terminal domain; ATPase family associated with various cellular activities (AAA); ATPase, AAA-type, core	Não	-
	ATP-dependent DNA ligase	evm.model.KI545856.1.179	map03030	ATP dependent DNA ligase domain; DNA ligase, ATP-dependent, central; ATP dependent DNA ligase C terminal region; DNA ligase, ATP-dependent, C-terminal; DNA ligase N terminus	Não	-
	DNA-directed DNA polymerase, family B, exonuclease domain	evm.model.KI545856.1.33	map03030	DNA Polymerase alpha zinc finger; Zinc finger, DNA-directed DNA polymerase, family B, alpha; DNA polymerase family B; DNA-directed DNA polymerase, family B, multifunctional domain; DNA polymerase alpha subunit p180 N terminal; DNA polymerase alpha catalytic subunit, N-terminal domain; DNA polymerase family B, exonuclease domain; DNA-directed DNA polymerase, family B,		

				exonuclease domain		
	Replication factor-A	evm.model.KI545858.1.103	map03030	Replication factor-A C terminal domain;; Replication factor A, C-terminal; Replication factor-A protein 1, N-terminal domain; Replication factor-A protein 1, N-terminal; OB-fold nucleic acid binding domain; OB-fold nucleic acid binding domain, AA-tRNA synthetase-type	Não	-
	Histone-like transcription factor/CCAAT	evm.model.KI545858.1.46	map03030	Histone-like transcription factor (CBF/NF-Y) and archaeal histone	Não	-
	DNA primase, large subunit	evm.model.KI545861.1.132	map03030	Eukaryotic and archaeal DNA primase, large subunit	Não	-
	DNA polymerase subunit Cdc27	evm.model.KI545861.1.45	map03030	DNA polymerase subunit Cdc27	Não	-
	DNA polymerase family B, exonuclease domain	evm.model.KI545862.1.15	map03030	DNA polymerase family B, exonuclease domain; DNA-directed DNA polymerase, family B, exonuclease domain; DNA polymerase family B; DNA-directed DNA polymerase, family B, multifunctional domain; C4-type zinc-finger of DNA polymerase delta; C4-type zinc-finger of DNA polymerase delta	Não	-
	Ribonuclease HII/HIII	evm.model.KI545863.1.144	map03030	Ribonuclease HII/HIII domain	Não	-
	PCNA (Proliferating cell nuclear antigen (cyclin))	evm.model.KI545864.1.133	map03030	Proliferating cell nuclear antigen, C-terminal domain; Proliferating cell nuclear antigen, PCNA, C-terminal; Proliferating cell nuclear antigen, C-terminal domain; Proliferating cell nuclear antigen, PCNA, C-terminal; Proliferating cell nuclear antigen, N-terminal domain; Proliferating cell nuclear antigen, PCNA, N-terminal	Sim	XLAIG
	RNase H	evm.model.KI545865.1.25	map03030	RNase H; Ribonuclease H domain	Não	-

	RNase H	evm.model.KI545865.1.27	map03030	RNase H; Ribonuclease H domain; Caulimovirus viroplasmin	Não	-
	DNA polymerase family B	evm.model.KI545873.1.409	map03030	DNA polymerase family B; DNA-directed DNA polymerase, family B, multifunctional domain; Domain of unknown function; DNA polymerase epsilon, catalytic subunit A, C-terminal; DNA polymerase family B, exonuclease domain; DNA-directed DNA polymerase, family B, exonuclease domain	Sim	XLAIG
	DNA polymerase delta, subunit 4	evm.model.KI545873.1.497	map03030	DNA polymerase delta, subunit	Não	-
	Replication factor C	evm.model.KI545873.1.709	map03030	ATPase family associated with various cellular activities (AAA); Replication factor C C-terminal domain; Replication factor C	Não	-
	endonuclease	evm.model.KI545884.1.108	map03030	XPG	Não	-
	DNA ligase	evm.model.KI545884.1.39	map03030	ATP dependent DNA ligase	Não	-
	DNA polymerase alpha/epsilon, subunit B	evm.model.KI545891.1.250	map03030	DNA polymerase alpha/epsilon subunit B	Sim	XLAIG
	DNA polymerase alpha/epsilon subunit B	evm.model.KI545891.1.42	map03030	DNA polymerase alpha/epsilon subunit B	Não	-

Via de Pentoses Fosfato (KEGG: map00030)

Gene em <i>K. brasiliensis</i> GHG001	Anotação	DEG	Group
evm.model.KI545851.1.504	Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain I (PF02878) Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain III (PF02880) Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain II (PF02879) Phosphoglucomutase/phosphomannomutase, C-terminal domain (PF00408)	Não	-

evm.model.KI545854.1.117	Transketolase, pyrimidine binding domain (PF02779) Transketolase, thiamine diphosphate binding domain (PF00456) Transketolase, C-terminal domain (PF02780)	Não	-
evm.model.KI545854.1.156	Glucose-6-phosphate dehydrogenase, NAD binding domain (PF00479) Glucose-6-phosphate dehydrogenase, C-terminal domain (PF02781)	Não	-
evm.model.KI545855.1.166	Ribose 5-phosphate isomerase A (phosphoriboisomerase A) (PF06026)	Não	-
evm.model.KI545855.1.178	Transaldolase (PF00923)	Não	-
evm.model.KI545856.1.66	pfkB family carbohydrate kinase (PF00294)	Não	-
evm.model.KI545857.1.98	Ribulose-phosphate 3 epimerase family (PF00834)	Não	-
evm.model.KI545859.1.163	SMP-30/Gluconolacunase/LRE-like region (PF08450)	Sim	XLOIG
evm.model.KI545861.1.90	Glucosamine-6-phosphate isomerasases/6-phosphogluconolactonase (PF01182)	Não	-
evm.model.KI545862.1.554	Membrane-bound protein with signal peptide	Sim	GLURG
evm.model.KI545862.1.555	NAD binding domain of 6-phosphogluconate dehydrogenase (PF03446) NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase (PF14833)	Sim	XLOIG
evm.model.KI545863.1.97	Phosphoribosyl synthetase-associated domain (PF14572) N-terminal domain of ribose phosphate pyrophosphokinase (PF13793)	Não	-
evm.model.KI545864.1.190	Transketolase, C-terminal domain (PF02780) Transketolase, thiamine diphosphate binding domain (PF00456) Transketolase, pyrimidine binding domain (PF02779)	Não	-
evm.model.KI545866.1.6	Transaldolase (PF00923)	Sim	XLOIG
evm.model.KI545873.1.218	DeoC/LacD family aldolase	Não	-
evm.model.KI545873.1.654	Fructose-bisphosphate aldolase class-II	Não	-
evm.model.KI545884.1.77	Ribose/Galactose Isomerase	Não	-
evm.model.KI545891.1.103	Fructose-1-6-bisphosphatase	Não	-
evm.model.KI545891.1.220	NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase; NAD binding domain of 6-phosphogluconate dehydrogenase; 6-phosphogluconate dehydrogenase, NADP-binding	Não	-
evm.model.KI545891.1.234	NAD binding domain of 6-phosphogluconate dehydrogenase; 6-phosphogluconate dehydrogenase, NADP-binding; 6-phosphogluconate dehydrogenase, C-terminal domain	Não	-

evm.model.KI545891.1.302	Phosphofructokinase	Não	-
evm.model.KI545893.1.6	Transaldolase (PF00923)	Não	-
evm.model.KI545893.1.81	N-terminal domain of ribose phosphate pyrophosphokinase (PF13793) Phosphoribosyl synthetase-associated domain (PF14572)	Não	-
evm.model.KI545894.1.161	NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase (PF14833) NAD binding domain of 6-phosphogluconate dehydrogenase (PF03446)	Não	-
evm.model.KI545894.1.267	Fructose-bisphosphate aldolase class-II (PF01116)	Sim	XLOIG
evm.model.KI545895.1.107	Shikimate kinase (PF01202)	Não	-
evm.model.KI545895.1.169	Phosphoglucose isomerase (PF00342)	Não	-

Ciclo do Ácido Cítrico (TCA) (KEGG: map00020)

Gene em <i>K. brasiliensis</i> GHG001	Anotação	DEG	Grupo
evm.model.KI545851.1.280	Biotin-requiring enzyme	Não	-
evm.model.KI545851.1.384	ATP-grasp domain; CoA-ligase	Não	-
evm.model.KI545851.1.421	lactate/malate dehydrogenase, alpha/beta C-terminal domain; lactate/malate dehydrogenase, NAD binding domain	Não	-
evm.model.KI545851.1.634	2Fe-2S iron-sulfur cluster binding domain; 4Fe-4S dicluster domain	Sim	XLORG
evm.model.KI545853.1.151	Transketolase, pyrimidine binding domain; Dehydrogenase E1 component	Não	-
evm.model.KI545854.1.166	lactate/malate dehydrogenase, NAD binding domain; lactate/malate dehydrogenase, alpha/beta C-terminal domain	Não	-
evm.model.KI545858.1.128	Phosphoenolpyruvate carboxykinase	Não	-
evm.model.KI545861.1.125	CybS	Sim	XLORG
evm.model.KI545861.1.134	Dehydrogenase E1 component	Não	-
evm.model.KI545862.1.185	Citrate synthase	Não	-
evm.model.KI545862.1.232	CoA binding domain; CoA-ligase	Não	-
evm.model.KI545862.1.472	CoA binding domain; CoA-ligase	Não	-
evm.model.KI545862.1.82	Biotin-requiring enzyme; 2-oxoacid dehydrogenases acyltransferase (catalytic domain)	Não	-
evm.model.KI545863.1.120	Transketolase, pyrimidine	Não	-

	binding domain; Transketolase, C-terminal domain		
evm.model.KI545863.1.126	Isocitrate/isopropylmalate dehydrogenase	Não	-
evm.model.KI545873.1.244	Fumarate reductase flavoprotein C-term; FAD binding domain	Sim	XLORG
evm.model.KI545873.1.384	HMGL-like; Conserved carboxylase domain; Biotin carboxylase C-terminal domain; Carbamoyl-phosphate synthase L chain, N-terminal domain; Biotin-requiring enzyme; Carbamoyl-phosphate synthase L chain, ATP binding domain	Não	-
evm.model.KI545873.1.412	Citrate synthase; CoA binding domain; CoA-ligase; ATP-grasp domain	Não	-
evm.model.KI545873.1.567	Biotin-requiring enzyme; e3 binding domain; 2-oxoacid dehydrogenases acyltransferase (catalytic domain)	Não	-
evm.model.KI545873.1.770	Lyase; Fumarase C C-terminus	Não	-
evm.model.KI545873.1.90	Isocitrate/isopropylmalate dehydrogenase	Não	-
evm.model.KI545873.1.91	Isocitrate/isopropylmalate dehydrogenase	Não	-
evm.model.KI545884.1.352	Pyridine nucleotide-disulphide oxidoreductase, dimerisation domain; Pyridine nucleotide-disulphide oxidoreductase;	Não	-
evm.model.KI545891.1.111	Aconitase C-terminal domain; Aconitase family (aconitate hydratase)	Não	-
evm.model.KI545892.1.71	Aconitase family (aconitate hydratase); Aconitase C-terminal domain	Sim	XLAIG
evm.model.KI545895.1.138	Succinate dehydrogenase/Fumarate reductase transmembrane subunit	Sim	XLORG

Fosforilação Oxidativa (KEGG: map00190)

Gene em <i>K. brasiliensis</i> GHG001	Anotação	DEG	Grupo
evm.model.KI545851.1.183	Cytochrome c oxidase subunit VIIc	Não	-
evm.model.KI545851.1.399	NADH(P)-binding	Não	-
evm.model.KI545851.1.528	Ubiquinol-cytochrome C reductase hinge protein	Não	-
evm.model.KI545851.1.634	2Fe-2S iron-sulfur cluster binding domain; 4Fe-4S dicluster domain	Sim	XLORG
evm.model.KI545853.1.29	NADH ubiquinone oxidoreductase subunit NDUFA12		

evm.model.KI545853.1.39	Vacuolar (H+)-ATPase G subunit	Não	-
evm.model.KI545854.1.161	Respiratory-chain NADH dehydrogenase, 49 Kd subunit	Não	-
evm.model.KI545855.1.131	NADH-ubiquinone oxidoreductase B15 subunit (NDUFB4)	Não	-
evm.model.KI545855.1.152	ATP synthase subunit D	Não	-
evm.model.KI545855.1.72	Ubiquinol-cytochrome C reductase complex 14kD subunit	Não	-
evm.model.KI545856.1.137	Cytochrome C1 family	Sim	XLORG
evm.model.KI545857.1.104	Cytochrome c oxidase subunit Vb	Não	-
evm.model.KI545857.1.123	Domain of unknown function (DUF1982); 2Fe-2S iron-sulfur cluster binding domain; NADH-ubiquinone oxidoreductase-G iron-sulfur binding region; Molybdopterin oxidoreductase	Sim	XLAIG
evm.model.KI545857.1.21	V-ATPase subunit C	Não	-
evm.model.KI545858.1.119	Mitochondrial F1-F0 ATP synthase subunit F of fungi	Não	-
evm.model.KI545858.1.159	NADH-ubiquinone oxidoreductase B12 subunit family	Não	-
evm.model.KI545858.1.167	ATP synthase	Não	-
evm.model.KI545859.1.12	ATP synthase j chain	Não	-
evm.model.KI545859.1.157	NADH dehydrogenase [ubiquinone] (complex I), alpha subcomplex, subunit	Não	-
evm.model.KI545859.1.21	Cytochrome c oxidase subunit IV	Não	-
evm.model.KI545859.1.59	V-ATPase subunit H	Não	-
evm.model.KI545860.1.159	Complex1_LYR-like	Não	-
evm.model.KI545860.1.22	Zinc-finger domain	Não	-
evm.model.KI545861.1.104	Rieske [2Fe-2S] domain; Ubiquinol cytochrome reductase transmembrane region	Sim	XLORG
evm.model.KI545861.1.125	CybS	Sim	XLORG
evm.model.KI545862.1.126	Mitochondrial ribosomal protein L51 / S25 / CI-B8 domain	Não	-
evm.model.KI545862.1.177	ATP synthase alpha/beta family, nucleotide-binding domain; ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain; ATP synthase alpha/beta chain, C terminal domain; ATPase, F1/V1/A1 complex, alpha/beta subunit, C-terminal; ATP synthase alpha/beta family, beta-barrel domain; ATPase, F1 complex alpha/beta subunit, N-terminal domain	Não	-
evm.model.KI545862.1.242	Respiratory-chain NADH dehydrogenase, 30 Kd subunit	Não	-
evm.model.KI545862.1.41	Insulinase (Peptidase family M16); Peptidase M16, N-terminal; Peptidase M16 inactive domain	Não	-
evm.model.KI545863.1.140	Ubiquinol-cytochrome C reductase, UQCRX/QCR9 like	Não	-
evm.model.KI545863.1.69	-	Não	-
evm.model.KI545863.1.89	Inorganic pyrophosphatase	Não	-
evm.model.KI545864.1.113	ATP synthase alpha/beta family, nucleotide-binding domain; ATPase, F1/V1/A1 complex, alpha/beta subunit, nucleotide-binding domain;	Não	-

	ATP synthase alpha/beta family, beta-barrel domain; ATPase, F1 complex alpha/beta subunit, N-terminal domain; ATP synthase alpha/beta chain, C terminal domain		
evm.model.KI545864.1.26	Mitochondrial ATP synthase B chain precursor (ATP-synt_B)	Não	-
evm.model.KI545864.1.65	-	Não	-
evm.model.KI545865.1.64	ATP synthase delta (OSCP) subunit	Não	-
evm.model.KI545866.1.12	ATP synthase D chain, mitochondrial (ATP5H)	Não	-
evm.model.KI545866.1.62	Cytochrome oxidase assembly protein Region of a membrane-bound protein predicted to be embedded in the membrane. Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm. Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	Não	-
evm.model.KI545867.1.42	Pyridine nucleotide-disulphide oxidoreductase; EF hand	Não	-
evm.model.KI545868.1.27	NADH-ubiquinone oxidoreductase B18 subunit (NDUFB7)	Não	-
evm.model.KI545868.1.95	VTC domain; SPX, N-terminal; Domain of unknown function (DUF202)	Não	-
evm.model.KI545869.1.1	Cytochrome C and Quinol oxidase polypeptide I	Não	-
evm.model.KI545869.1.2	NADH-Ubiquinone/plastoquinone (complex I), various chains	Não	-
evm.model.KI545873.1.107	GRIM-19 protein	Não	-
evm.model.KI545873.1.213	Cation transporter/ATPase, N-terminus; Cation-transporting P-type ATPase, N-terminal; haloacid dehalogenase-like hydrolase; HAD-like domain; E1-E2 ATPase	Não	-
evm.model.KI545873.1.244	Fumarate reductase flavoprotein C-term; FAD binding domain	Sim	XLORG
evm.model.KI545873.1.275	Cytochrome c oxidase subunit VIa	Não	-
evm.model.KI545873.1.317	ATP synthase, Delta/Epsilon chain, beta-sandwich domain	Não	-
evm.model.KI545873.1.444	Mitochondrial ATP synthase g subunit	Não	-
evm.model.KI545873.1.476	ETC complex I subunit conserved region	Não	-
evm.model.KI545873.1.480	ATP synthase subunit C	Não	-
evm.model.KI545873.1.595	ATP synthase alpha/beta chain, C terminal domain; ATP synthase alpha/beta family, beta-barrel domain; ATP synthase alpha/beta family, nucleotide-binding domain	Não	-
evm.model.KI545873.1.606	V-type ATPase 116kDa subunit family	Não	-
evm.model.KI545873.1.609	4Fe-4S dicluster domain	Sim	XLAIG
evm.model.KI545873.1.610	Thioredoxin-like [2Fe-2S] ferredoxin	Sim	XLAIG
evm.model.KI545873.1.697	SLBB domain; Respiratory-chain NADH dehydrogenase 51 Kd subunit; NADH-ubiquinone oxidoreductase-F iron-sulfur binding region	Sim	XLAIG
evm.model.KI545873.1.706	Cytochrome c oxidase assembly protein	Não	-

	CtaG/Cox11		
evm.model.KI545873.1.759	Phosphopantetheine attachment site	Não	-
evm.model.KI545884.1.118	ATP synthase subunit C	Não	-
evm.model.KI545884.1.247	ATP synthase complex subunit h	Não	-
evm.model.KI545884.1.312	Complex 1 protein (LYR family)	Não	-
evm.model.KI545884.1.366	Cytochrome c oxidase subunit Va	Não	-
evm.model.KI545884.1.99	NADH ubiquinone oxidoreductase, 20 Kd subunit	Sim	XLAIG
evm.model.KI545891.1.166	Cytochrome oxidase c subunit VIb	Não	-
evm.model.KI545891.1.17	ATP synthase (C/AC39) subunit	Não	-
evm.model.KI545891.1.219	ATP synthase subunit H	Não	-
evm.model.KI545891.1.230	Cation transporter/ATPase, N-terminus; haloacid dehalogenase-like hydrolase; E1-E2 ATPase	Não	-
evm.model.KI545892.1.258	ATP synthase E chain	Não	-
evm.model.KI545892.1.285	NADH ubiquinone oxidoreductase subunit NDUFA12	Não	-
evm.model.KI545893.1.197	ATP synthase alpha/beta chain, C terminal domain; ATP synthase alpha/beta family, beta-barrel domain; ATP synthase alpha/beta family, nucleotide-binding domain	Não	-
evm.model.KI545894.1.34	ATP synthase (F/14-kDa) subunit	Não	-
evm.model.KI545894.1.6	Cytochrome oxidase c subunit VIb	Não	-
evm.model.KI545895.1.136	Cytochrome C oxidase copper chaperone (COX17)	Não	-
evm.model.KI545895.1.138	Succinate dehydrogenase/Fumarate reductase transmembrane subunit	Sim	XLORG
evm.model.KI545895.1.48	UcrQ family	Não	-
evm.model.KI545895.1.86	ATP synthase (E/31 kDa) subunit	Não	-
evm.model.KI545895.1.92	NADH-ubiquinone oxidoreductase ASHI subunit (CI-ASHI or NDUFB8)	Não	-

Via de Interconversão de Pentose and Glucuronato (KEGG: map00040)

Gene	Annotation	DEG	Group
evm.model.KI545851.1.130	UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain (PF03720) UDP-glucose/GDP-mannose dehydrogenase family, central domain (PF00984) UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain (PF03721)	Não	-
evm.model.KI545851.1.191	Oxidoreductase family, NAD-binding Rossmann fold (PF01408)	Não	-
evm.model.KI545851.1.430	Zinc-binding dehydrogenase (PF00107) Alcohol dehydrogenase GroES-like domain (PF08240)	Não	-
evm.model.KI545851.1.438	Alcohol dehydrogenase GroES-like domain (PF08240) Zinc-binding dehydrogenase (PF00107)	Não	-
evm.model.KI545852.1.102	Pectinesterase (PF01095) NADH(P)-binding (PF13460)	Não	-
evm.model.KI545852.1.108	Aldehyde dehydrogenase family (PF00171)	Não	-

evm.model.KI545857.1.98	Ribulose-phosphate 3 epimerase family (PF00834)	Não	-
evm.model.KI545859.1.143	UTP--glucose-1-phosphate uridylyltransferase (PF01704)	Não	-
evm.model.KI545862.1.425	Zinc-binding dehydrogenase (PF00107) Alcohol dehydrogenase GroES-like domain (PF08240)	Não	-
evm.model.KI545891.1.306	Aldehyde dehydrogenase family (PF00171)	Sim	XLAIG
evm.model.KI545891.1.37	Aldehyde dehydrogenase family (PF00171)	Não	-
evm.model.KI545894.1.122 *	Zinc-binding dehydrogenase (PF00107) Alcohol dehydrogenase GroES-like domain (PF08240)	Sim	GLURG
evm.model.KI545894.1.214	short chain dehydrogenase (PF00106)	Não	-
evm.model.KI545895.1.131 *	Aldo/keto reductase family (PF00248)	Não	-

* Caracterização descrita na dissertação de mestrado de Borges (2014).

Metabolismo de Frutose e Manose (KEGG: map00051)

Gene em <i>K. brasiliensis</i> GHG001	Anotação	DEG	Grupo
evm.model.KI545851.1.121	DAK2 domain; Dak1 domain	Não	-
evm.model.KI545851.1.430	Zinc-binding dehydrogenase (PF00107) Alcohol dehydrogenase GroES-like domain (PF08240)	Não	-
evm.model.KI545851.1.438	Alcohol dehydrogenase GroES-like domain (PF08240) Zinc-binding dehydrogenase (PF00107)	Não	-
evm.model.KI545855.1.166	Ribose 5-phosphate isomerase A (phosphoriboisomerase A) (PF06026)	Não	-
evm.model.KI545857.1.8	Eukaryotic phosphomannomutase	Não	-
evm.model.KI545862.1.161	Membrane-bound protein with signal peptide	Não	-
evm.model.KI545864.1.3	Enoyl-(Acyl carrier protein) reductase	Sim	GLURG
evm.model.KI545866.1.5	Mannose-6-phosphate isomerase, Firmicutes type, short form	Sim	XLOIG
evm.model.KI545868.1.24	Phosphomannose isomerase type I	Sim	XLOIG
evm.model.KI545873.1.402	6-phosphofructo-2-kinase; Histidine phosphatase superfamily (branch 1)	Não	-
evm.model.KI545873.1.654	Fructose-bisphosphate aldolase class-II	Não	-
evm.model.KI545884.1.77	Ribose/Galactose Isomerase	Não	-
evm.model.KI545891.1.103	Fructose-1-6-bisphosphatase	Não	-
evm.model.KI545891.1.191	6-phosphofructo-2-kinase	Não	-
evm.model.KI545891.1.302	Phosphofructokinase domain	Não	-
evm.model.KI545892.1.119	Mannose-6-phosphate isomerase	Sim	GLUIG
evm.model.KI545892.1.278	Hexokinase	Não	-
evm.model.KI545892.1.292	Triosephosphate isomerase	Não	-
evm.model.KI545893.1.88	Triosephosphate isomerase	Não	-
evm.model.KI545894.1.144	Hexokinase	Não	-
evm.model.KI545894.1.267	Fructose-bisphosphate aldolase class-II	Sim	XLOIG
evm.model.KI545894.1.86	Enolase C-terminal domain-like; Mandelate racemase / muconate lactonizing enzyme, C-terminal domain	Sim	XLOIG

Metabolismo de Glutationa (KEGG: map00480)

Gene em <i>K. brasiliensis</i> GHG001	Anotação	DEG	Grupo
evm.model.KI545851.1.437	Glutathione S-transferase, C-terminal domain	Não	-
evm.model.KI545851.1.89	NAD dependent epimerase/dehydratase family; Hydantoinase B/oxoprolinase	Não	-
evm.model.KI545853.1.15	Ribonucleotide reductase, barrel domain; ATP cone domain; Ribonucleotide reductase, all-alpha domain	Sim	XLAIG
evm.model.KI545854.1.156	Glucose-6-phosphate dehydrogenase, NAD binding domain; Glucose-6-phosphate dehydrogenase, C-terminal domain	Não	-
evm.model.KI545854.1.82	Carboxymuconolactone decarboxylase family	Não	-
evm.model.KI545857.1.152	AhpC/TSA family	Não	-
evm.model.KI545857.1.25	Glutathione S-transferase, N-terminal domain	Não	-
evm.model.KI545859.1.3	Carboxymuconolactone decarboxylase family	Não	-
evm.model.KI545859.1.4	Methyltransferase domain	Sim	XLOIG
evm.model.KI545859.1.60	Eukaryotic glutathione synthase; Eukaryotic glutathione synthase, ATP binding domain	Não	-
evm.model.KI545860.1.98	DSBA-like thioredoxin domain	Não	-
evm.model.KI545862.1.24	Glutathione S-transferase, C-terminal domain; Glutathione S-transferase, N-terminal domain	Não	-
evm.model.KI545862.1.345	Glutathione peroxidase	Não	-
evm.model.KI545862.1.554	Membrane-bound protein with signal peptide	Sim	GLURG
evm.model.KI545862.1.555	NAD binding domain of 6-phosphogluconate dehydrogenase; NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase	Sim	XLOIG
evm.model.KI545863.1.126	Isocitrate/isopropylmalate dehydrogenase	Não	-
evm.model.KI545863.1.8	Alpha/beta hydrolase family; Carboxymuconolactone decarboxylase family (ver anotação de CAZymes)	Sim	XLOIG
evm.model.KI545866.1.50	C-terminal domain of 1-Cys peroxiredoxin; AhpC/TSA family	Não	-
evm.model.KI545866.1.75	Ribonucleotide reductase, small chain	Sim	GLURG
evm.model.KI545873.1.282	Pyridine nucleotide-disulphide oxidoreductase	Não	-
evm.model.KI545873.1.380	Pyridoxal-dependent decarboxylase, C-terminal sheet domain; Pyridoxal-dependent decarboxylase, pyridoxal binding domain	Sim	GLUIG
evm.model.KI545884.1.173	Gamma-glutamyltranspeptidase	Não	-
evm.model.KI545884.1.17	Spermine/spermidine synthase	Não	-
evm.model.KI545891.1.220	NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase; NAD binding domain of 6-phosphogluconate dehydrogenase; 6-phosphogluconate dehydrogenase, NADP-binding	Não	-
evm.model.KI545891.1.234	NAD binding domain of 6-phosphogluconate dehydrogenase; 6-phosphogluconate dehydrogenase, NADP-binding; 6-phosphogluconate dehydrogenase, C-terminal domain	Não	-
evm.model.KI545892.1.126	Redoxin	Não	-
evm.model.KI545893.1.246	Redoxin	Não	-
evm.model.KI545893.1.79	Cytosol aminopeptidase family, catalytic domain; Cytosol aminopeptidase family, N-terminal domain	Não	-
evm.model.KI545894.1.130	Peptidase family M20/M25/M40; Peptidase dimerisation domain	Não	-
evm.model.KI545894.1.161	NAD-binding of NADP-dependent 3-hydroxyisobutyrate dehydrogenase (PF14833)		

	NAD binding domain of 6-phosphogluconate dehydrogenase (PF03446)		
evm.model.KI545894.1.182	Gamma-glutamyltranspeptidase	Não	-
evm.model.KI545894.1.92	Gamma-glutamyltranspeptidase	Não	-
evm.model.KI545895.1.143	Glutamate-cysteine ligase	Não	-
evm.model.KI545895.1.219	Glutathione S-transferase, N-terminal domain	Não	-
evm.model.KI545895.1.34	Hydantoinase/oxoprolinase N-terminal region; Hydantoinase/oxoprolinase N-terminal region	Não	-

Metabolismo de açúcar de amino e nucleotídeo (KEGG: map00520)

Gene em <i>K. brasiliensis</i> GHG001	Anotação	DEG	Grupo
evm.model.KI545851.1.130	UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain (PF03720) UDP-glucose/GDP-mannose dehydrogenase family, central domain (PF00984) UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain (PF03721)	Não	-
evm.model.KI545851.1.480	Chitin synthase	Não	-
evm.model.KI545851.1.492	Chitin synthase N-terminal; Chitin synthase	Não	-
evm.model.KI545851.1.504	Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain I; Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain III; Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain II; Phosphoglucomutase/phosphomannomutase, C-terminal domain	Não	-
evm.model.KI545851.1.643	Alpha-L-arabinofuranosidase C-terminus; Carbohydrate binding domain (ver anotação de CAZymes)	Sim	XLAIG
evm.model.KI545854.1.80	Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain II; Phosphoglucomutase/phosphomannomutase, alpha/beta/alpha domain I; Phosphoglucomutase/phosphomannomutase, C-terminal domain	Não	-
evm.model.KI545855.1.18	Chitin synthase N-terminal; Chitin synthase	Não	-
evm.model.KI545855.1.38	Glycosyl hydrolases family 18; (Anotação de CAZymes: GH18); Possible functions (www.cazy.org): chitinase (EC 3.2.1.14); lysozyme (EC 3.2.1.17); endo-β-N-acetylglucosaminidase (EC 3.2.1.96); peptidoglycan hydrolase with endo-β-N-acetylglucosaminidase specificity (EC 3.2.1.-); Nod factor hydrolase (EC 3.2.1.-); xylanase inhibitor; concanavalin B; narbonin	Sim	XLAIG
evm.model.KI545857.1.8	Eukaryotic phosphomannomutase	Não	-
evm.model.KI545859.1.143	UTP--glucose-1-phosphate uridylyltransferase	Não	-
evm.model.KI545859.1.37	Chitin synthase	Não	-
evm.model.KI545860.1.152	Chitin synthase N-terminal; Chitin synthase	Sim	GLUIG
evm.model.KI545860.1.19	Polysaccharide deacetylase	Não	-
evm.model.KI545861.1.57	Oxidoreductase FAD-binding domain	Não	-

evm.model.KI545862.1.278	Amidohydrolase	Não	-
evm.model.KI545862.1.281	Glucosamine-6-phosphate isomerases/6-phosphogluconolactonase	Sim	XLOIG
evm.model.KI545862.1.33	Alpha-L-arabinofuranosidase B, catalytic (ver anotação de CAZymes)	Sim	XLAIG
evm.model.KI545862.1.392	RNA polymerase II transcription elongation factor	Não	-
evm.model.KI545862.1.393	Carbohydrate binding domain; Alpha-L-arabinofuranosidase C-terminus (ver anotação de CAZymes)	Não	-
evm.model.KI545862.1.442	Galactose-1-phosphate uridyl transferase, N-terminal domain; Galactose-1-phosphate uridyl transferase, C-terminal domain	Não	-
evm.model.KI545862.1.541	Galactokinase galactose-binding signature; GHMP kinases C terminal; GHMP kinases N terminal domain	Não	-
evm.model.KI545863.1.130	Chitin synthase	Não	-
evm.model.KI545863.1.75	NAD dependent epimerase/dehydratase family; UDP-glucose 4-epimerase C-term subunit	Sim	XLOIG
evm.model.KI545864.1.162	Alpha-L-arabinofuranosidase C-terminus (ver anotação de CAZymes)	Sim	XLOIG
evm.model.KI545864.1.182	Glycosyl hydrolase family 62 (ver anotação de CAZymes)	Sim	XLAIG
evm.model.KI545866.1.28	Glycosyl hydrolase family 3 N terminal domain; Glycosyl hydrolase family 3 C-terminal domain; Fibronectin type III-like domain (ver anotação de CAZymes)	Sim	XLAIG
evm.model.KI545866.1.5	Mannose-6-phosphate isomerase, Firmicutes type, short form	Sim	XLOIG
evm.model.KI545868.1.24	Phosphomannose isomerase type I	Sim	XLOIG
evm.model.KI545868.1.44	Glycosyl hydrolases family 18 (ver anotação de CAZymes)	Não	-
evm.model.KI545873.1.622	Oxidoreductase NAD-binding domain; Oxidoreductase FAD-binding domain	Não	-
evm.model.KI545873.1.676	Glycosyl hydrolase family 20, catalytic domain; beta-acetyl hexosaminidase like (ver anotação de CAZymes)	Não	-
evm.model.KI545884.1.256	Membrane-bound protein	Não	-
evm.model.KI545891.1.114	Polysaccharide deacetylase (ver anotação de CAZymes)	Sim	XLARG
evm.model.KI545891.1.241	Chitin synthase; Cytochrome b5-like Heme/Steroid binding domain	Não	-
evm.model.KI545892.1.119	Mannose-6-phosphate isomerase	Sim	GLUIG
evm.model.KI545892.1.231	DEK C terminal domain; Cytochrome b5-like Heme/Steroid binding domain; Chitin synthase	Sim	XLARG
evm.model.KI545892.1.278	Hexokinase	Não	-
evm.model.KI545892.1.66	Acetyltransferase (GNAT) family	Não	-
evm.model.KI545892.1.78	UTP--glucose-1-phosphate uridylyltransferase	Não	-
evm.model.KI545892.1.9	NAD dependent epimerase/dehydratase family	Não	-
evm.model.KI545893.1.184	Cytochrome b5-like Heme/Steroid binding domain; Myosin head (motor domain); Chitin synthase; DEK C terminal domain	Não	-
evm.model.KI545893.1.189	Glutamine amidotransferases class-II; SIS domain	Não	-
evm.model.KI545894.1.144	Hexokinase	Não	-
evm.model.KI545894.1.260	Acetyltransferase (GNAT) family	Não	-
evm.model.KI545895.1.169	Phosphoglucose isomerase (PF00342)	Não	-

Tabela adicional 5 - Basidiomicetos (Agaricomycotina e Pucciniomycotina) incluídos na filogenia usando os quatro marcadores descritos em Capella-Gutierrez et al. (2014).

Classe	Espécie	Degradação de Biomassa **	Fonte	Data
Agaricomycotina	<i>Phanerochaete chrysosporium</i> (Pch)	Degradação branca	JGI	14/12/2015
Agaricomycotina	<i>Trametes versicolor</i> (Tvr)	Degradação branca	JGI	14/12/2015
Agaricomycotina	<i>Schizophyllum commune</i> (Scm)	Degradação parda / branca	JGI	14/12/2015
Agaricomycotina	<i>Schizophyllum commune</i> H4-8 (Scm)	Degradação parda / branca	JGI	14/12/2015
Agaricomycotina	<i>Agaricus bisporus</i> (Abi)	Resíduos de lixo	JGI	14/12/2015
Agaricomycotina	<i>Coprinopsis cinerea</i> (Cci)	Resíduos de lixo	JGI	14/12/2015
Agaricomycotina	<i>Laccaria bicolor</i> (Lbi)	Não	JGI	14/12/2015
Agaricomycotina	<i>Fomitopsis pinicola</i> (Fpi)	Degradação parda	JGI	14/12/2015
Agaricomycotina	<i>Ceriporiopsis subvermispora</i> (Gsa)	Degradação branca	JGI	14/12/2015
Agaricomycotina	<i>Malassezia globosa</i> (Mgl)		JGI	14/12/2015
Agaricomycotina	<i>Coniophora puteana</i> (Cpt)	Degradação parda	JGI	14/12/2015
Agaricomycotina	<i>Wolfiporia cocos</i> (Wco)	Degradação parda	JGI	14/12/2015
Agaricomycotina	<i>Gloeophyllum trabeum</i> (Gtr)	Degradação parda	JGI	14/12/2015
Agaricomycotina	<i>Dichomitus squalens</i> (Dsq)	Degradação branca	JGI	14/12/2015
Agaricomycotina	<i>Wallemia sebi</i> (Wse)		JGI	14/12/2015
Agaricomycotina	<i>Punctularia strigosozonata</i> (Psa)	Degradação branca	JGI	14/12/2015
Agaricomycotina	<i>Fomitiporia mediterranea</i> (Fma)	Degradação branca	JGI	14/12/2015
Agaricomycotina	<i>Cryptococcus neoformans</i> (Cne)	Não	JGI	14/12/2015
Pucciniomycotina	<i>Rhodotorula graminis</i> (Rgr)	-	JGI	14/12/2015
Pucciniomycotina	<i>Melampsora larici-populina</i> (Mlr)	-	JGI	14/12/2015
Pucciniomycotina	<i>Puccinia triticina</i> (Pta)	-	Broad Institute	14/12/2015
Pucciniomycotina	<i>Mixia osmundae</i>	-	JGI	14/12/2015

a	(Mos)			
Pucciniomycotina	<i>Puccinia graminis</i>	-	JGI	14/12/2015

* JGI - proteomas de Basidiomicetos foram baixados da plataforma Mycocosm do Joint Genome Institute (JGI)

** Referências: Lundell et al., 2014; Busk e Lange, 2015

Tabela adicional 6 - Macrovariáveis associadas aos genomas usados nas análises comparativas (*Ustilaginaceae* + *M. globosa*).

Organismo	Origem	Ecologia	Fitopatógeno	Patógeno <i>H. sapiens</i>
<i>Anthracocystis flocculosa</i> PF-1	Planta			
<i>K. brasiliensis</i> GHG001	Inseto (trato intestinal de larva)	Saprotrófico, simbótico, or desconhecido	Assume-se que não	
<i>Moesziomyces antarcticus</i> JCM10317			Assume-se que não	Sim (oportunista)
<i>Moesziomyces antarcticus</i> T-34	Planta (exsudato)		Assume-se que não	Sim (oportunista)
<i>Moesziomyces aphidis</i> DSM70725	Planta (superfície de folha, mas em secreção de inseto)		Assume-se que não	Sim (oportunista)
<i>Moesziomyces</i> sp. F16C1 *	Planta (flor)		Assume-se que não	
<i>Moesziomyces</i> sp. F5C1 *	Planta (flor)		Assume-se que não	
<i>Pseudozyma hubeiensis</i> SY62			Assume-se que não	
<i>Pseudozyma</i> sp. F8B2 *	Planta (flor)		Assume-se que não	
<i>Sporisorium reilianum</i> SRZ2		Biotrófico	Sim (milho)	
<i>Sporisorium scitamineum</i> 2014001		Biotrófico	Sim (cana-de-açúcar)	
<i>Sporisorium scitamineum</i> CSIRO		Biotrófico	Sim (cana-de-açúcar)	
<i>Sporisorium scitamineum</i> SSC39		Biotrófico	Sim (cana-de-açúcar)	
<i>Ustilago esculenta</i> MMT			Sim (“Manchurian wild rice”)	
<i>Ustilago hordei</i> Uh4857-4		Biotrófico	Sim (cevada)	
<i>Mycosarcoma maydis</i> 521		Biotrófico	Sim (milho)	
<i>Malassezia globosa</i> **		Patógeno animal não obrigatório (Xu et al., 2007)	Não	Sim

* Genomas Sequenciados neste trabalho

** Grupo externo das análises filogenéticas

Tabela adicional 7 - Resultados de expansão de contração de famílias de proteínas (CAZymes e

TAPs) em *Ustilaginaceae*.

Família	Nó / Clado (Newick)	Tamanho da família (Nó)	P valor (CAFE)	P valor ajustado
C2C2-GATA	(MAP1,MAN1)	9	0,000139716	0,002235456
SET	MAP1	9	0,000212574	0,003401184
HB	MF82	4	0,000361486	0,005783776
AA7	(SSC1,SSC3)	3	0,000376775	0,0060284
Top	MAP1	13	0,000402561	0,006440976
GH16	(SSC2,(SSC1,SSC3))	16	0,000461637	0,007386192
Fungal-trans2	SSC1	2	0,000797391	0,012758256
GH43	SSC2	2	0,000837193	0,013395088
AA1	SSC2	2	0,000837193	0,013395088
MYB	UMA1	3	0,000903282	0,014452512
GH28	(MAP1,MAN1)	1	0,00139376	0,0148667733
GH28	MF16	2	0,00114213	0,0148667733
GH23	(MAP1,MAN1)	1	0,00139376	0,02230016
GH78	(MAP1,MAN1)	1	0,00139376	0,02230016
GH93	MAP1	2	0,00139376	0,02230016
GH26	(MAP1,MAN1)	1	0,00139376	0,02230016
AA1	((SSC2,(SSC1,SSC3)),SRM1)	2	0,00272549	0,02230016
AA1	(MAP1,MAN1)	2	0,00278752	0,02230016

GH28	((((SSC2,(SSC1,SSC3)),SRM1),(UMA1,(MF82,PHU1))),KBR1)	2	0,00317786	0,02542288
Fungal-trans2	(SSC2,(SSC1,SSC3))	1	0,00238963	0,0254893867
C1-4-t1	SSC2	0	0,00250878	0,02676032
C1-4-t1	(SSC2,(SSC1,SSC3))	1	0,00238963	0,02676032
CBM35	SSC2	0	0,00250878	0,02676032
CBM35	((((SSC2,(SSC1,SSC3)),SRM1),(UMA1,(MF82,PHU1))),KBR1)	1	0,00219545	0,02676032
AA7	SSC2	4	0,00251157	0,02679008
MYB	(MAP1,MAN1)	2	0,00278752	0,0297335467
C1-4-t1	MAP1	0	0,00417354	0,03338832
GH27	MF16	3	0,00228427	0,03654832
GH43	(MF16,MF51)	6	0,0034944	0,0372736
GH115	SSC1	0	0,00238963	0,03823408
CBM41	SSC1	0	0,00238963	0,03823408
CE14	(SSC2,(SSC1,SSC3))	1	0,00238963	0,03823408
GT28	SSC1	0	0,00238963	0,03823408
GNAT	MF82	14	0,00274059	0,04384944

Tabela adicional 8 - Número de diferenças (substituições, inserções e deleções) em comparações das regiões ITS1, ITS2, 5.8S e subunidade maior ribossomal entre as leveduras sequenciadas neste trabalho e as mais próximas na análise filogenética (figura adicional 2).

		Número de diferenças								
		ITS1	ITS 1	ITS2	ITS 2	5.8 S	5.8S	LS U	LS U	
Organismo	Organismo comparação	Sbt	indel	Sbt	indel	Sbt	indel	Sbt	indel	

<i>Moesziomyces sp.</i> F16C1	<i>Moesziomyces aphidis</i>	1	5	5	12	0	0	3	0
<i>Moesziomyces sp.</i> F16C1	<i>Moesziomyces antarcticus</i>	3	4	4	8	0	0	1	0
<i>Moesziomyces sp.</i> F16C1	<i>Moesziomyces rugulosus</i>	4	3	6	14	0	0	3	0
<i>Moesziomyces sp.</i> F16C1	<i>Moesziomyces sp. F5C1</i>	0	0	0	0	0	0	0	0
<i>Moesziomyces sp. F5C1</i>	<i>Moesziomyces aphidis</i>	1	5	5	12	0	0	3	0
<i>Moesziomyces sp. F5C1</i>	<i>Moesziomyces antarcticus</i>	3	5	4	8	0	0	1	0
<i>Moesziomyces sp. F5C1</i>	<i>Moesziomyces rugulosus</i>	4	3	6	14	0	0	3	0
<i>Pseudozyma sp. F8B2</i>	<i>Pseudozyma hubeiensis</i>	5	6	3	0	0	0	0	0
<i>Pseudozyma sp. F8B2</i>	<i>Macalpinomyces arundinellae-setosae</i>	5	12	2	7	0	0	Nd	Nd

Sbt = substituições; Nd = Não disponível; Indel = inserções/deleções.

Anexo 3 - Lista de publicações relacionadas ao mestrado

2015 (a partir de agosto)

- dos Santos, Renato Augusto Corrêa, et al. "Draft genome sequence of *Komagataeibacter intermedius* strain AF2, a producer of cellulose, isolated from Kombucha tea." *Genome announcements* 3.6 (2015): e01404-15.

Resumo: Here, we present the draft genome sequence of *Komagataeibacter intermedius* strain AF2, which was isolated from Kombucha tea and is capable of producing cellulose, although at lower levels compared to another bacterium from the same environment, *K. rhaeticus* strain AF1.

2016

- Brown, Neil Andrew, et al. "RNAseq reveals hydrophobins that are involved in the adaptation of *Aspergillus nidulans* to lignocellulose." *Biotechnology for biofuels* 9.1 (2016): 145.

Resumo: Background: Sugarcane is one of the world's most profitable crops. Waste steam-exploded sugarcane bagasse (SEB) is a cheap, abundant, and renewable lignocellulosic feedstock for the next-generation biofuels. In nature, fungi seldom

exist as planktonic cells, similar to those found in the nutrient-rich environment created within an industrial fermenter. Instead, fungi predominantly form biofilms that allow them to thrive in hostile environments.

Results: In turn, we adopted an RNA-sequencing approach to interrogate how the model fungus, *Aspergillus nidulans*, adapts to SEB, revealing the induction of carbon starvation responses and the lignocellulolytic machinery, in addition to morphological adaptations. Genetic analyses showed the importance of hydrophobins for growth on SEB. The major hydrophobin, RodA, was retained within the fungal biofilm on SEB fibres. The StuA transcription factor that regulates fungal morphology was up-regulated during growth on SEB and controlled hydrophobin gene induction. The absence of the RodA or DewC hydrophobins reduced biofilm formation. The loss of a RodA or a functional StuA reduced the retention of the hydrolytic enzymes within the vicinity of the fungus. Hence, hydrophobins promote biofilm formation on SEB, and may enhance lignocellulose utilisation via promoting a compact substrate-enzyme-fungus structure.

Conclusion: This novel study highlights the importance of hydrophobins to the formation of biofilms and the efficient deconstruction of lignocellulose.

2017

1. de Vries, Ronald P., et al. "Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*." *Genome biology* 18.1 (2017): 28.

Resumo: Background: The fungal genus *Aspergillus* is of critical importance to humankind. Species include those with industrial applications, important pathogens of humans, animals and crops, a source of potent carcinogenic contaminants of food, and an important genetic model. The genome sequences of eight aspergilli have already been explored to investigate aspects of fungal biology, raising questions about evolution and specialization within this genus.

Results: We have generated genome sequences for ten novel, highly diverse *Aspergillus* species and compared these in detail to sister and more distant genera. Comparative studies of key aspects of fungal biology, including primary and secondary metabolism, stress response, biomass degradation, and signal transduction, revealed both conservation and diversity among the species. Observed

genomic differences were validated with experimental studies. This revealed several highlights, such as the potential for sex in asexual species, organic acid production genes being a key feature of black aspergilli, alternative approaches for degrading plant biomass, and indications for the genetic basis of stress response. A genome-wide phylogenetic analysis demonstrated in detail the relationship of the newly genome sequenced species with other aspergilli.

Conclusions: Many aspects of biological differences between fungal species cannot be explained by current knowledge obtained from genome sequences. The comparative genomics and experimental study, presented here, allows for the first time a genus-wide view of the biological diversity of the aspergilli and in many, but not all, cases linked genome differences to phenotype. Insights gained could be exploited for biotechnological and medical applications of fungi.

2. Dos Reis, Thaila Fernanda, et al. "The low affinity glucose transporter HxtB is also involved in glucose signalling and metabolism in *Aspergillus nidulans*." *Scientific Reports* 7 (2017): 45073.

Resumo: One of the drawbacks during second-generation biofuel production from plant lignocellulosic biomass is the accumulation of glucose, the preferred carbon source of microorganisms, which causes the repression of hydrolytic enzyme secretion by industrially relevant filamentous fungi. Glucose sensing, subsequent transport and cellular signalling pathways have been barely elucidated in these organisms. This study therefore characterized the transcriptional response of the filamentous fungus *Aspergillus nidulans* to the presence of high and low glucose concentrations under continuous chemostat cultivation with the aim to identify novel factors involved in glucose sensing and signalling. Several transcription factor- and transporter-encoding genes were identified as being differentially regulated, including the previously characterized glucose and xylose transporter HxtB. HxtB was confirmed to be a low affinity glucose transporter, localizing to the plasma membrane under low- and high-glucose conditions. Furthermore, HxtB was shown to be involved in conidiation-related processes and may play a role in downstream glucose signalling. A gene predicted to encode the protein kinase PskA was also identified as being important for glucose metabolism. This study identified several proteins with predicted roles in glucose metabolic processes and provides a foundation for further

investigation into the response of biotechnologically important filamentous fungi to glucose.

3. Augusto Corrêa dos Santos, Renato, Gustavo Henrique Goldman, and Diego Mauricio Riaño-Pachón. "ploidyNGS: visually exploring ploidy with Next Generation Sequencing data." *Bioinformatics* (2017): btx204.

Resumo: ploidyNGS is a model-free, open source tool to visualize and explore ploidy levels in a newly sequenced genome, exploiting short read data. We tested ploidyNGS using both simulated and real NGS data of the model yeast *Saccharomyces cerevisiae*. ploidyNGS allows the identification of the ploidy level of a newly sequenced genome in a visual way.

4. Borin, Gustavo Pagotto, et al. "Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by *Aspergillus niger* and *Trichoderma reesei*." *BMC genomics* 18.1 (2017): 501.

Resumo: Background: Second generation (2G) ethanol is produced by breaking down lignocellulosic biomass into fermentable sugars. In Brazil, sugarcane bagasse has been proposed as the lignocellulosic residue for this biofuel production. The enzymatic cocktails for the degradation of biomass-derived polysaccharides are mostly produced by fungi, such as *Aspergillus niger* and *Trichoderma reesei*. However, it is not yet fully understood how these microorganisms degrade plant biomass. In order to identify transcriptomic changes during steam-exploded bagasse (SEB) breakdown, we conducted a RNA-seq comparative transcriptome profiling of both fungi growing on SEB as carbon source.

Results: Particular attention was focused on CAZymes, sugar transporters, transcription factors (TFs) and other proteins related to lignocellulose degradation. Although genes coding for the main enzymes involved in biomass deconstruction were expressed by both fungal strains since the beginning of the growth in SEB, significant differences were found in their expression profiles. The expression of these enzymes is mainly regulated at the transcription level, and *A. niger* and *T. reesei* also showed differences in TFs content and in their expression. Several sugar transporters that were induced in both fungal strains could be new players on

biomass degradation besides their role in sugar uptake. Interestingly, our findings revealed that in both strains several genes that code for proteins of unknown function and pro-oxidant, antioxidant, and detoxification enzymes were induced during growth in SEB as carbon source, but their specific roles on lignocellulose degradation remain to be elucidated.

Conclusions: This is the first report of a time-course experiment monitoring the degradation of pretreated bagasse by two important fungi using the RNA-seq technology. It was possible to identify a set of genes that might be applied in several biotechnology fields. The data suggest that these two microorganisms employ different strategies for biomass breakdown. This knowledge can be exploited for the rational design of enzymatic cocktails and 2G ethanol production improvement.

Anexo 4 - Códigos usados no trabalho

1. Código em R para análise de enriquecimento de mapas do KEGG, CAZymes e TAPs.

```
# Author: Dr. rer. nat. Diego Mauricio Riaño Pachón
computeFisherExactTest <-function(listGenes,geneGroups,sp) {
  EnrichmentResults<-data.frame(stringsAsFactors = FALSE)
  #geneGroups must be a dataframe with at least two columns, the first
  #column MUST be the geneID and the second the groupName (family, kegg, etc)
  #listGenes must be a dataframe with at least one column, i.e., the
  #selected geneIDs
  for (grp in unique(geneGroups[,2])) {
    GROUP_LIST_tbl<-unique(as.character(geneGroups[ which(geneGroups[,1]
    %in% listGenes[,1] & geneGroups[,2] == grp),1]))
    GROUP_noLIST_tbl<-unique(as.character(geneGroups[
    which(!(geneGroups[,1] %in% listGenes[,1]) & geneGroups[,2] == grp),1)))
    noGROUP_LIST_tbl<-unique(as.character(geneGroups[
    which(geneGroups[,1] %in% listGenes[,1] & geneGroups[,2] != grp),1)))
    noGROUP_noLIST_tbl<-unique(as.character(geneGroups[
    which(!(geneGroups[,1] %in% listGenes[,1]) & geneGroups[,2] != grp),1)))

    GROUP_LIST_cnt<-length(GROUP_LIST_tbl)
    GROUP_noLIST_cnt<-length(GROUP_noLIST_tbl)
    noGROUP_LIST_cnt<-length(noGROUP_LIST_tbl)
    noGROUP_noLIST_cnt<-length(noGROUP_noLIST_tbl)

    myMatrix<-
    matrix(c(GROUP_LIST_cnt, GROUP_noLIST_cnt, noGROUP_LIST_cnt, noGROUP_noLIST_cnt,
    nrow = 2,
    dimnames=
      list(c("LIST","noLIST"),
           c(grp,"noGROUP")))
    res<-fisher.test(myMatrix)
    if(GROUP_LIST_cnt>0 & GROUP_noLIST_cnt>0) {
      EnrichmentResults<-
      rbind(EnrichmentResults,data.frame(cbind(grp,sp, GROUP_LIST_cnt, GROUP_noLIST
```

```

    _cnt,noGROUP_LIST_cnt,noGROUP_noLIST_cnt,as.numeric(res$p.value),as.numeric
    (res$estimate)))
  }
}
colnames(EnrichmentResults)<-
c("Family","Species","GROUP_LIST_cnt","GROUP_noLIST_cnt","noGROUP_LIST_cnt"
,"noGROUP_noLIST_cnt","Pvalue","OddsRatio")
  EnrichmentResults$Pvalue_adjusted<-p.adjust(p =
as.numeric(as.character(EnrichmentResults$Pvalue)), method = 'fdr')
  return(EnrichmentResults)
}

```

Anexo 5 - Participação em eventos e apresentação de trabalhos

2016

Participação em congresso com apresentação de trabalho: 62º Congresso Brasileiro de Genética. Título do trabalho: “Phylogenomic Analysis and Comparative Genomics of *Ustilaginaceae* (Basidiomycetes): Exploiting Evolutionary Information to Uncover Probable Regulatory Mechanisms of CAZYme Expression”.

Curso Teórico de Perícia Criminal: Genética Forense.

Minicurso: Quantificação de DNA nuclear e cromossômica por Citometria de Fluxo e de Imagem.

2017

Workshop: Essential Statistics with R (University of Virginia)

II Workshop of Systems Microbiology.

Workshop: Power and Sample Size Analysis (University of Virginia)

Participação em congresso com apresentação de trabalho: Brazilian BioEnergy Science and Technology Conference 2017. Título do trabalho: “Understanding the metabolism of hemicellulosic sugars from Omics of the yeast-like fungus *Kalmanozyma brasiliensis* GHG001 and the Comparative Genomics with *Ustilaginaceae* (Basidiomycetes)”.

Anexo 6 - Contribuição da bolsa de estágio e pesquisa no exterior (BEPE-FAPESP)

Apesar das diversas dificuldades associadas com a reconstrução metabólica de eucariotos, em particular de organismos pouco estudados e que não apresentam muitos dados experimentais disponíveis, durante o desenvolvimento do projeto de mestrado foi reconstruído o rascunho do modelo de redes metabólicas de *K. brasiliensis* GHG001. O trabalho foi realizado durante estágio de pesquisa no

exterior (FAPESP nº 2016/13968-1) no Computational Systems Biology Laboratory, Universidade de Virginia e supervisionado pelo Dr. Jason Papin.

O rascunho do modelo metabólico gerado está disponível no repositório online: https://github.com/SantosRAC/UNICAMP_RACSMaster. Nele, também estão disponíveis os scripts usados para a reconstrução. A tabela adicional 9 apresenta estatísticas gerais dos modelos de *K. brasiliensis* GHG001, antes e depois de melhoramento e curadoria realizados durante o estágio no exterior. Além disso, estão apresentadas estatísticas do modelo de *M. maydis* 521, gentilmente fornecido pelo doutorando Christian Lieven e ao Dr. Thiemo Zambanini e que foi usado para melhorar o modelo de *K. brasiliensis*.

Tabela adicional 9 - Número de componentes (genes, reações, metabólitos) nos modelos metabólicos usados neste trabalho.

	Modelo original de <i>K. brasiliensis</i> GHG001 *	<i>M. maydis</i> 521	Modelo misto **	Modelo curado
Número de genes	760	1.079	-	994
Número de reações	868	2.218	1.713	1.510
Número de metabólitos	764	1.987	1.550	1.476
Número de precursores de biomassa produzidos sem reações de troca no modelo ***	-	-	40	34
Número de “infeasible loops” ****	-	-	420	308

* Modelo de *K. brasiliensis* GHG001 gerado automaticamente usando métodos computacionais no PRIAM.

** Modelo de *K. brasiliensis* GHG001 após inclusão de reações e metabólitos de *M. maydis* 521, realizada após busca por ortólogos descrita nas análises de genômica comparativa.

*** O modelo ainda não apresenta reações de troca de massa com o ambiente (entrada e saída); assim, não é possível que resultados de simulações apresentem geração de precursores da reação de produção de biomassa (crescimento); para que seja funcional, o modelo não deve apresentar produção de nenhum dos precursores em simulações sem reações de troca com o ambiente.

**** “Loops” impossíveis do ponto de vista termodinâmico.

As principais conclusões sobre o desenvolvimento do modelo:

1. O modelo metabólico ainda precisa ser melhorado para que possa ser funcional.

2. Os principais pontos que merecem atenção para melhoramento do modelo incluem:

A. padronização de identificadores de genes, reações, compartimentos celulares, metabólitos, no modelo metabólico (atualmente no formato SBML3) e geração de tabelas com um formato que possam ser facilmente editadas por pessoas envolvidas em experimentos (geralmente sem conhecimento sobre o processamento computacional dos modelos). Uma solução seria a geração de modelos a partir de dados tabulados com SBTAB (<http://sbtab.net/>).

B. Inclusão de informações de extensões no modelo em formato SBML que permitam a realização de simulações, como simulações de Análise de Balanço de Fluxo (*Flux Balance Analysis*, FBA);

C. Realização de experimentos para resolver ambiguidades introduzidas no modelo pela ausência de informação sobre o uso de cofatores por muitas enzimas e de composição de biomassa em *K. brasiliensis*.

D. Curadoria de “loops” internos no modelo termodinamicamente impossíveis, que envolve a correção de direcionalidade das reações;

E. Curadoria de reações que geram precursores de biomassa sem reações de troca de massa com o ambiente (geração de massa “do além”);

Devido à importância de modelos metabólicos em projetos de Biologia de Sistemas, os resultados obtidos durante o estágio de pesquisa no exterior pelo aluno de mestrado poderão servir como ponto de partida para novos projetos envolvendo o melhoramento do modelo metabólico de *K. brasiliensis*.

Anexo 7 - Iniciativas educacionais

2015

Grupo de Estudos em Python (CTBE, CNPEM) (informal)

2016

Grupo de Estudos em Biologia Molecular e Bioinformática (UFSCar, Campus Araras/SP)

Workshop de Python para Raspagem de Dados

Contribuição para a aula de Genômica (Flávia V. Winck, IQ, USP)

Grupo de Estudos em Python (CTBE, CNPEM) (informal)

2017

Instrutor do Workshop de Python para Dados Biológicos (Instituto de Química, Univ. de São Paulo)

DECLARAÇÃO BIOÉTICA E BIOSSEGURANÇA



COORDENADORIA DE PÓS-GRADUAÇÃO
INSTITUTO DE BIOLOGIA
Universidade Estadual de Campinas
Caixa Postal 6109, 13083-970, Campinas, SP, Brasil
Fone (19) 3521-6378. email: cpgib@unicamp.br



DECLARAÇÃO

Em observância ao §5º do Artigo 1º da Informação CCPG-UNICAMP/001/15, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Dissertação de Mestrado, intitulada "*Análise genômica da levedura xilanolítica Pseudozyma brasiliensis GHG001*", desenvolvida no Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura: _____
Nome do(a) aluno(a): Renato Augusto Corrêa dos Santos

Assinatura: _____
Nome do(a) orientador(a): Gustavo Henrique Goldman

Data: 28 de fevereiro de 2018

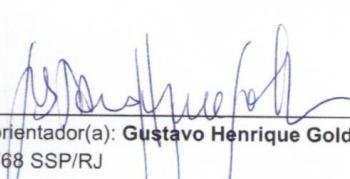
DECLARAÇÃO DE DIREITOS AUTORAIS

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Análise genômica da levedura xilanolítica Pseudozyma brasiliensis GHG001**, não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 28 de fevereiro de 2018

Assinatura : 
Nome do(a) autor(a): **Renato Augusto Corrêa dos Santos**
RG n.º 471550644 SSP/SP

Assinatura : 
Nome do(a) orientador(a): **Gustavo Henrique Goldman**
RG n.º 4824768 SSP/RJ