

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Departamento de Genética, Ecologia e Evolução
Programa de Pós-Graduação em Bioinformática

Aline Soares dos Reis

A ASSINATURA GENÔMICA DE FUNGOS NOS NÍVEIS DE GENES E VIAS

Belo Horizonte

2021

Aline Soares dos Reis

A ASSINATURA GENÔMICA DE FUNGOS NOS NÍVEIS DE GENES E VIAS

Dissertação apresentada ao Programa de Pós-Graduação em Bioinformática do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito para obtenção do título de Mestre em Bioinformática.

Orientador: Professor Doutor Francisco Pereira Lobo

Belo Horizonte

2021

043 Reis, Aline Soares dos.
 A assinatura genômica de fungos nos níveis de genes e vias [manuscrito] /
 Aline Soares dos Reis. – 2021.
 86 f. : il. ; 29,5 cm.

 Orientador: Professor Doutor Francisco Pereira Lobo.
 Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto
 de Ciências Biológicas. Programa Interunidades de Pós-Graduação em
 Bioinformática.

 1. Bioinformática. 2. Genômica. 3. Fungos. I. Lobo, Francisco Pereira . II.
 Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III.
 Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa Interunidades de Pós-Graduação em Bioinformática da UFMG

ATA DA DEFESA DE DISSERTAÇÃO

ALINE SOARES DOS REIS

Às quatorze horas do dia **31 de maio de 2021**, reuniu-se, no aplicativo Zoom, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**A ASSINATURA GENÔMICA DE FUNGOS NOS NÍVEIS DE GENES E VIAS**", requisito para obtenção do grau de Mestre em **Bioinformática**. Abrindo a sessão, o Presidente da Comissão, **Dr. Francisco Pereira Lobo**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Professor/Pesquisador	Instituição	Indicação
Dr. Francisco Pereira Lobo	Universidade Federal de Minas Gerais	Aprovada
Dr. Aristóteles Góes Neto	Universidade Federal de Minas Gerais	Aprovada
Dr. Heron Oliveira Hilário	Universidade Federal de Minas Gerais	Aprovada

Pelas indicações, a candidata foi considerada: **Aprovada**

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 31 de maio de 2021.



Documento assinado eletronicamente por **Heron Oliveira Hilário, Usuário Externo**, em 31/05/2021, às 17:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

Documento assinado eletronicamente por **Aristoteles Goes Neto, Professor do Magistério**



Superior, em 31/05/2021, às 17:27, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Francisco Pereira Lobo, Professor do Magistério Superior**, em 10/06/2021, às 11:41, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site

[https://sei.ufmg.br/sei/controlador_externo.php?](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)

[acao=documento_conferir&id_orgao_acesso_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0753810** e o código CRC **3E7691FE**.

AGRADECIMENTOS

A Deus, por ter sido tudo possível, mesmo quando as chances pareciam improváveis.

Aos meus familiares, a quem dedico essa alegria e conquista, especialmente meus pais Jane e Fausto e minha vizinha Penha (*in memoriam*), que mesmo sem ler e escrever foi a melhor professora da minha vida. Também agradeço aos meus irmãos Samuel e Davi, minha avó Bernadete, meus padrinhos e tios Adilson, Dulcinéia, Fátima e Maurício. Vocês me ensinam todos os dias o que é amor incondicional, humildade, simplicidade, fé e esperança. Em cada um de vocês eu encontro apoio e inspiração para ser cada dia um ser humano melhor e lutar pelos meus sonhos, sem esquecer minhas origens.

Aos meus amigos, em especial Pedro, Lucy, Pâmella, Ramon e Camilla, com os quais pude dividir alegrias e desalentos. Obrigada pela força, suporte e paciência.

Aos professores que tive ao longo da vida, especialmente no Programa de Pós-Graduação em Bioinformática. Agradeço carinhosamente também à professora Glória e aos secretários Tiago e Sheila, pela paciência e suporte nos momentos de dúvidas e dificuldades.

Aos meus colegas de jornada, os “Franciscanos” e amigos do Laboratório de Algoritmos em Biologia (LAB), especialmente Alison, Agnello, Giovanni e Igor, por fazerem dos meus dias mais alegres e por terem me ajudado inúmeras vezes a entender mais sobre a ciência e a vida.

Por fim, ao meu orientador Chico, que se tornou um dos pilares da minha formação e exemplo de profissional. Com quem aprendi que ser cientista é sempre ter o que aprender, a questionar certezas e ser cada dia um pouco melhor que o dia anterior. Obrigada por me aceitar como aluna, pelo suporte em tantos momentos, por toda paciência, compreensão e, principalmente, pelos ensinamentos que me ajudaram a ampliar minha visão de mundo, dentro e fora do ambiente acadêmico.

*“Se enxerguei mais longe é
porque me apoiei sobre ombros
de gigantes” Isaac Newton*

RESUMO

Os fungos são um grupo monofilético de organismos eucarióticos que possuem grande diversidade bioquímica, fisiológica e morfológica extrema. Esses organismos desempenham um importante papel ecológico como decompositores e também têm grande importância econômica para a produção de bebidas, alimentos, produtos farmacêuticos, biocombustíveis e outros processos biotecnológicos. Por outro lado, várias espécies de fungos também são grandes ameaças à saúde e à agricultura humanas como patógenos de seres humanos, outros animais e plantas. O uso não racional e generalizado de agentes antifúngicos é uma preocupação relevante nesse contexto, dado o risco de seleção cepas de fungos resistentes. Outra preocupação é a toxicidade de uma fração considerável desses medicamentos antifúngicos para os hospedeiros devido à semelhança molecular do hospedeiro e dos parasitas, já que ambos são organismos eucarióticos e, conseqüentemente, mais próximos evolutivamente do que eucariotos e bactérias, por exemplo. Portanto, os fungos são um grupo interessante para caracterizar em nível molecular e funcional a fim de compreender melhor o *modus operandi* celular desse táxon e também detectar possíveis novos alvos para intervenção molecular em fungos patogênicos. O *Kyoto Gene and Genome Encyclopedia* (KEGG) é um banco de dados que fornece informações selecionadas sobre grupos de genes homólogos entre espécies, definidas como grupos Kegg Orthology (KO). O KEGG também aloca KO em camadas de funções biológicas de ordem superior, fornecendo as funções enzimáticas dos grupos KO em diferentes vias bioquímicas curadas, apresentando assim uma camada comum de componentes genômicos e suas funções entre as espécies. Neste trabalho, usamos uma estratégia genômica comparativa de duas etapas (nível de genes e vias) para procurar papéis enzimáticos e módulos biológicos significativamente enriquecidos em genomas de fungos em comparação com outros organismos eucarióticos. Usamos a API KEGG para obter os 535 organismos eucarióticos encontrados nesse banco de dados em janeiro de 2020, juntamente com sua classificação taxonômica. Usamos essas informações para dividir esses genomas eucarióticos em 129 grupos de fungos (F) e 406 não-fungos (NF). Também usamos a API KEGG para baixar os grupos KO observados em cada genoma eucariótico, bem como suas descrições. Para buscar KO significativamente enriquecido no grupo F, utilizamos o seguinte algoritmo: 1) realizar bootstrap nos genomas nos dois grupos 100 vezes, obtendo 100 pseudo-réplicas do mesmo tamanho que os dados originais; 2) Para cada inicialização, execute o teste exato de Fisher, seguido da correção de FDR, para encontrar significativamente mais KOs representados em F em comparação com NF; 3) um KO foi considerado super-representado em F se fosse super-representado (valor $q < 0,05$) em pelo menos 95% das repetições e ausente no grupo NF. Do total de 13962 KOs encontrados em pelo menos um genoma eucariótico, descobrimos que 495 são significativamente enriquecidos no grupo F. Entre os 50 KOs mais abundantes, encontramos marcas claras de suas funções ecológicas como decompositores (por exemplo, K19356

- *lytic cellulose monooxygenase (C1-hydroxylating)* [EC:1.14.99.54]) e suas atividades metabólicas de interesse comercial (por exemplo, K19305-NP11; *deuterolysin* [EC: 3.4.24.39]). Curiosamente, também encontramos várias enzimas já direcionadas por medicamentos antifúngicos (por exemplo, K23787 - *DMA1_2*; *E3 ubiquitin-protein ligase DMA1/2* [CE: 2.3.2.23] e K19564 - *CTR, HNM1; choline transport protein*), que indica que nossa estratégia também encontrou alvos de medicamentos já rotulados como interessantes por outros pesquisadores. Finalmente, encontramos vários candidatos interessantes para pesquisas futuras que desempenham um papel importante no estilo de vida dos fungos, como reguladores de transcrição (por exemplo, K09043 - *YAP; AP-1-like transcription factor*) e componentes de processos de parede celular (por exemplo, K13682 - *BMT, WRY; beta-1,2-mannosyltransferase* [EC:2.4.1.-]). Para realizar uma análise genômica comparativa no nível de via, começamos usando a API KEGG para obter, para cada rota descrita neste banco de dados, todos os seus KOs. Segundo, procuramos todas as vias de KO enriquecido nos genomas dos fungos, classificando todas as vias por razão do número de KOs enriquecidos na via para a contagem total de KO. Entre as 15 vias que contêm as maiores frações de KOs de fungos enriquecidos, 7 são vias de KEGG específicas para fungos (por exemplo, map04113 - *Meiosis - yeast* and map04011 - *MAPK signaling pathway - yeast*). Também encontramos vias filogeneticamente diversas que contêm altas frações de KO enriquecidas em genomas de fungos, como reparo de DNA (map03450 - *Non-homologous end-joining*), síntese de proteínas (map03008 - *Ribosome biogenesis in eukaryotes*) e metabolismo de ácidos graxos (map01040 - *Biosynthesis of unsaturated fatty acids*), que podem compreender objetivos interessantes para uma intervenção mais sistêmica na biologia de fungos (por exemplo, busca de *chokepoints*). Por fim, este trabalho destaca alvos já conhecidos dos medicamentos e candidatos interessantes para futuras pesquisas em biologia fúngica.

PALAVRAS-CHAVE: KEGG. Biologia de Fungos. Genômica Comparativa.

ABSTRACT

Fungi are a monophyletic group of eukaryotic organisms possessing extreme biochemical, physiological and morphological diversity. These organisms play an important ecological role as decomposers and also have major economic importance for the production of beverages, foods, pharmaceuticals, biofuels and other biotechnological processes. On the other hand, several fungi species are also major threats to human health and agriculture as pathogens of humans, other animals and plants. The widespread non-rational usage of antifungal agents is a major concern, as it increasingly selected resistant fungal strains. Another concern is the host toxicity of a considerable fraction of such drugs due to the molecular similarity of host and parasites, as both are closely related eukaryotic organisms. Therefore, fungi are an interesting group to be characterized at the molecular and functional levels for both having a better understanding of the cellular *modus operandi* of this taxon and also to detect potential new targets for molecular intervention in pathogenic fungi. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database that provides curated information about groups of homologous genes across species, defined as Kegg Ortholog (KO) groups. KEGG also maps KOs into higher-order layers of biological functions by providing the enzymatic functions of KO groups in different curated biochemical pathways, therefore providing a common layer of genomic components and their functional roles across species. In this work, we used a two-step comparative genomics strategy (gene- and pathway-level) to search for enzymatic roles and biological modules significantly enriched in fungal genomes when compared with other eukaryotic organisms. We used KEGG API to obtain all 535 eukaryotic organisms found in this database as in January 2020, together with their taxonomic classification. We used this information to split these eukaryotic genomes in 129 Fungi (F) and 406 non-Fungi (NF) groups. We also used KEGG API to download the KO groups observed in each eukaryotic genome, as well as their descriptions. To search for KOs significantly enriched in F group, we used the following algorithm: 1) bootstrap genomes in both groups 100 times, obtaining 100 pseudo-replicates of the same size as the original data; 2) for each bootstrap, execute Fisher's exact test, followed by FDR correction, to search for KOs significantly more represented in F when compared to NF; 3) a KO was considered over represented in F if it was found as over represented ($q\text{-value} < 0.05$) in at least 95% of the replicates, and absent in NF group. From the total of 13962 KOs found in at least one eukaryotic genome, we found 495 to be significantly enriched in F group. Among the 50 most abundant KOs, we found clear marks of their ecological roles as decomposers (e.g. K19356 - *lytic cellulose monooxygenase (C1-hydroxylating) [EC:1.14.99.54]*) and of their metabolic activities of commercial interest (e.g. K19305 - *NPII; deuterolysin [EC:3.4.24.39]*). Interestingly, we also found several enzymes already targeted by antifungal drugs (e.g. K23787 - *DMA1_2; E3 ubiquitin-protein ligase DMA1/2 [EC:2.3.2.23]* and K19564 - *CTR, HNM1; choline transport protein*), indicating our strategy also found drug targets already labeled as interesting by other researchers. Finally, we found several interesting candidates for future research that play major roles in fungal lifestyles, such as transcriptional regulators (e.g. K09043 - *YAP; AP-1-like transcription factor*) and components of cell wall processes (e.g. K13682 - *BMT, WRY; beta-1,2-mannosyltransferase [EC:2.4.1.-]*). To perform a pathway-level comparative genomics analysis, we started by using KEGG API to obtain, for each pathway described in this database, all its KOs. Secondly, we searched all pathways for the enriched KOs in fungal genomes, ranking all pathways by the ratio of the number of enriched KOs in the pathway by the total KO count. Among the 15 pathways containing the largest fractions of enriched fungal KOs, 7 are fungi-specific KEGG pathways (e.g. *map04113 - Meiosis - yeast* and *map04011 - MAPK signaling pathway - yeast*). We also found more phylogenetically diverse pathways containing high fractions of KOs enriched in fungal genomes, such as DNA repair (*map03450 - Non-homologous end-joining*), protein synthesis (*map03008 - Ribosome biogenesis in eukaryotes*) and fatty acid metabolism (*map01040 - Biosynthesis of unsaturated fatty acids*), which may comprise interesting targets for a more systemic intervention in fungal biology (e.g. search for *chokepoints*). Together, this work highlights both known drug targets and interesting candidates for future research of fungal biology.

KEYWORDS: KEGG. Fungal biology. Comparative genomics.

LISTA DE FIGURAS

Figura 1. Esquema dos tipos de homologia entre genes da proteína globina

Figura 2. Índice de Bancos de Dados KEGG

Figura 3. Representação KEGG da via da glicólise e gliconeogênese e página de informações

Figura 4. Esquema da integração dos bancos de dados KEGG por meio do KEGG Orthology

Figura 5. Demonstração gráfica dos tipos de reações de *chokepoints*

Figura 6. Workflow das etapas da obtenção dos dados deste projeto

Figura 7. Arquivo dos genomas contendo o ID, código e descrição de cada organismo contido no KEGG

Figura 8. Filtragem dos organismos utilizados para este estudo

Figura 9. Exemplo de um arquivo antes e depois da curadoria

Figura 10. Arquivo contendo os identificadores dos KOs dos eucariotos e suas respectivas descrições

Figura 11. Arquivo de parâmetros do CALANGO

Figura 12. Arquivos das vias metabólicas KEGG

Figura 13. Exemplo do conteúdo dos arquivos de vias KEGG gerados

Figura 14. Lista das vias metabólicas e a contagem de KOs enriquecidos em fungos

Figura 15. Exemplo de entrada no KEGG Mapper

Figura 16. Exemplo de saída do KEGG Mapper

Figura 17. KOs enriquecidos em fungos mapeados na via metabólica map04011 - *MAPK signaling pathway*

Figura 18. KOs enriquecidos em fungos mapeados na via metabólica map04111 - *Cell Cycle (Yeast)*

Figura 19. KOs enriquecidos em fungos mapeados na via metabólica map04113 - *Meiosis (Yeast)*

Figura 20. Via da Biossíntese de vários tipos de N-Glicanos com os KOs fúngicos enriquecidos em destaque.

LISTA DE TABELAS

Tabela 1. KOs enriquecidos ordenados por valor-q

Tabela 2. Vias metabólicas KEGG enriquecidas em KOs fúngicos – Ordenadas pela razão de KOs enriquecidos nos fungos

Tabela 3. Vias metabólicas KEGG enriquecidas em KOs fúngicos – Ordenadas pela quantidade decrescente de KOs enriquecidos nos fungos

LISTA DE ABREVIATURAS

KEGG – Kyoto Encyclopedia of Genes and Genomes

KO – Kegg Orthology

API – Application Programming Interface

REST – Representational State Transfer

F – Grupo de organismos Fungos

NF – Grupo de organismos Não-Fungos

EC – Enzyme Commission Number

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Fungos - Características gerais, origem, diversidade, papéis ecológicos e ambientais	12
1.2	Parasitismo em fungos	14
1.3	Genômica de fungos	15
1.4	Genômica comparativa	17
1.5	Conceitos de evolução de genes - homólogos, ortólogos, parálogos e xenólogos	18
1.6	Bancos de dados de vias bioquímicas - KEGG	19
1.6.1	KEGG Pathway	20
1.6.2	KEGG Orthology	22
1.6.3	KEGG API e Protocolo REST	23
1.7	Reações de <i>chokepoints</i>	24
2	OBJETIVOS	26
2.1	Objetivo geral	26
2.2	Objetivos específicos	26
3	MÉTODOS	27
3.1	Estrutura computacional utilizada neste projeto	27
3.2	Obtenção dos organismos presentes no KEGG e utilizados nesse estudo	27
3.3	Seleção de genomas de fungos e demais eucariotos disponíveis no KEGG	28
3.4	Obtenção dos grupos de homólogos (KOs) dos organismos	30
3.5	Curadoria dos arquivos (Identificadores + KOs)	31
3.6	Montagem do dicionário de KOs	33
3.7	Classificação dos fungos de acordo com a patogenicidade	34
3.8	Obtenção dos KOs significativamente enriquecidos nos genomas de fungos comparados aos demais eucariotos	35
3.8.1	Análises estatísticas por meio do CALANGO	35
3.8.2	Obtenção das vias metabólicas KEGG e associação aos KOs	36
3.8.3	Mapeamento dos KOs enriquecidos nas vias metabólicas KEGG	38
3.9	Busca por <i>chokepoints</i> nas vias KEGG com KOs fúngicos enriquecidos	39

4 RESULTADOS	42
4.1 KOs mais significativamente super-representados em fungos em relação aos eucariotos não-fungos	42
4.2 Vias metabólicas KEGG enriquecidas em KOs fúngicos	43
4.3 Candidatos a <i>chokepoints</i> nas vias KEGG contendo KOs fúngicos enriquecidos	47
5 DISCUSSÃO	49
5.1 KOs mais significativamente super-representados em fungos em relação aos eucariotos não-fungos	49
5.2 Vias metabólicas KEGG enriquecidas em KOs fúngicos	51
5.3 Candidatos a <i>chokepoints</i> nas vias KEGG contendo KOs fúngicos enriquecidos	53
6 CONCLUSÕES	55
7 PERSPECTIVAS	57
REFERÊNCIAS	58
MATERIAL SUPLEMENTAR	69

1 INTRODUÇÃO

1.1 - Fungos - Características gerais, origem, diversidade, papéis ecológicos e ambientais

A palavra fungo provém do latim *fungus* (cogumelo), e deriva da palavra grega *sphongos* (esponja) que faz referência à morfologia macroscópica dos cogumelos que são os representantes clássicos dos fungos, apesar de constituírem uma fração muito restrita do Reino Fungi. Já o termo micologia advém das junções das palavras gregas *mykes* (cogumelo) e *logos* (discurso) e se refere ao estudo sistemático dos fungos (PERSOON, 1796; SIMPSON, 1979).

O Reino Fungi é um grupo monofilético de organismos eucarióticos que possui alta diversidade bioquímica, fisiológica e morfológica. Assim, os fungos possuem tanto características comuns aos demais eucariotos como especificidades que os definem como um reino à parte. De modo geral, são organismos heterotróficos, que podem ser simbióticos ou de vida livre, uni ou pluricelulares e possuem uma parede celular de constituição exclusiva ao reino formada por quitina, glucanos e proteínas glicosiladas (HIBBETT et al., 2007; WATKINSON; BODDY; MONEY, 2015; MADIGAN et al., 2016; TORTORA, 2017).

A estrutura somática dos fungos pode ser classificada como leveduriforme (leveduras unicelulares) ou filamentosa (hifas multicelulares, também conhecidas como mofos ou bolores). Podem se reproduzir de forma assexuada ou sexuada, sendo a produção de esporos sua principal forma de propagação. Sua nutrição se dá por meio absorptivo, realizam respiração celular ou fermentação para obtenção de energia e possuem glicogênio como reserva energética (WATKINSON; BODDY; MONEY, 2015; MADIGAN et al., 2016; TORTORA, 2017).

Atualmente, os registros apontam que os fungos, animais e plantas se divergiram há aproximadamente 1 bilhão de anos e são os únicos três reinos que desenvolveram a multicelularidade em ambientes terrestres. Igualmente aos demais eucariotos, os fungos também são provenientes de ancestrais aquáticos. No entanto, existem poucos registros fósseis, o que dificultou o entendimento das origens dos fungos por muito tempo (BRUNS, 2006).

Antigamente, os padrões evolutivos dos fungos eram mal compreendidos devido à falta de dados moleculares disponíveis e, em função disso, por muito tempo

os fungos fizeram parte do reino vegetal em virtude de algumas semelhanças morfológicas e de habitat com as plantas (GALAGAN et al., 2005). No entanto, a partir do surgimento das ferramentas de análises moleculares, além da possibilidade de análise de diversas características microscópicas, logo se evidenciou que os fungos são mais próximos filogeneticamente dos animais do que dos vegetais, e que existem peculiaridades (como por exemplo, a presença e o tipo de parede celular) que os tornavam um reino separado dos demais eucariotos (BRUNS, 2006).

Atualmente, existe uma classificação aceita em que os fungos são distribuídos em 224 ordens dentro de 12 filos: Basidiomycota, Ascomycota, Microsporidia, Chytridiomycota, Zoopagomycota, Mucoromycota, Cryptomycota, Neocallimastigomycota, Entorrhizomycota, Aphelidiomycota, Monoblepharidomycota e Blastocladiomycota. Os dois grandes filos mais representativos do reino em espécies descritas são os filos Ascomycota (com cerca de 83.837 espécies) e Basidiomycota (com cerca de 48.405 espécies) (NARANJO-ORTIZ; GABALDÓN, 2019; JAMES et al., 2020).

A ampla diversidade dos fungos também é refletida pela distribuição global dos membros desse táxon, os quais estão presentes em todos os ecossistemas do mundo, seja na terra, no ar, na água, no solo, nos vegetais ou na microbiota natural de humanos e diversos animais. Isso significa que estes organismos adquiriram ao longo de sua história evolutiva a capacidade de adaptação aos diversos ambientes, com relatos de espécies que habitam ambientes extremos como o deserto do Atacama, salinas hipersalinas e em rochas da Antártica (POINTING et al., 2009; HAWKSWORTH, 2012; MOHANTA; BAE, 2015).

As interações ecológicas descritas nos fungos incluem saprofitismo, predação e diversas simbioses, como mutualismo, comensalismo ou parasitismo. Os fungos saprotróficos desempenham um papel ecológico fundamental sendo o maior grupo de fungos, bem como os principais decompositores de matéria orgânica do mundo, juntamente às bactérias. Por serem detentores de um amplo aparato de processos metabólicos e de biotransformação, são capazes de degradar inclusive materiais de difícil metabolismo por outros organismos, como lignocelulose e quitina, o que os torna essenciais no ciclo do carbono e em processos biotecnológicos (TUNLID; TALBOT, 2002; VÁRNAI et al., 2014; STRULLU-DERRIEN et al., 2018).

Devido a suas propriedades metabólicas, os fungos também possuem relevância econômica, pois fazem parte de diversos processos que envolvem os setores energético, agrônomo, alimentício, farmacêutico e técnico-científico. Exemplos da utilização destes organismos incluem a produção de bebidas e alimentos fermentados, antibióticos, biocombustíveis, utilização como biofábricas enzimáticas e na biorremediação. Os fungos também são importantes na área científica, pois algumas espécies são utilizadas como organismo modelo, por exemplo *Saccharomyces cerevisiae* (levedura) e *Neurospora crassa* (filamentoso) (BARKE et al., 2010; RINEAU et al., 2012; MOHANTA; BAE, 2015).

1.2 Parasitismo em fungos

Se por um lado os fungos trazem grandes benefícios econômicos, por outro as interações de parasitismo também caracterizam os fungos como importantes agentes patogênicos. Os fungos patogênicos podem ser encontrados tanto em hospedeiros do reino animal quanto do reino vegetal, o que traz grande impacto na agricultura e saúde humana e animal, sendo capazes de causar diversas condições patológicas (TUNLID; TALBOT, 2002). Diferentemente dos demais parasitas, os fungos são particularmente problemáticos em termos de tratamentos disponíveis devido, principalmente, às semelhanças biológicas que compartilham com os demais eucariotos (GALAGAN et al., 2005).

Vários fatores contribuem para um fungo patogênico. Fungos capazes de causar patologias em outros organismos possuem características moleculares, como capacidade de sobreviver e crescer no ambiente único dos corpos dos hospedeiros. A capacidade de resistir a temperaturas elevadas, a flutuações no pH, ao estresse oxidativo e a fatores imunológicos são exemplos de vantagens importantes que tornam vários tipos de fungos causadores de doenças em plantas e animais (CASADEVALL, 2018; SHERRINGTON et al., 2018).

O uso não racional generalizado de agentes antifúngicos também compreende um problema relevante nesse cenário, uma vez que pode selecionar cepas de fungos resistentes aos fármacos. Outra preocupação é a toxicidade relativamente alta de uma fração considerável das drogas antifúngicas para os hospedeiros (baixa toxicidade seletiva) devido à semelhança molecular entre

hospedeiro e parasitas, já que ambos são organismos eucarióticos (CASADEVALL, 2018; FUENTEFRIA, 2018).

1.3 Genômica de fungos

O sequenciamento do genoma do organismo modelo *Saccharomyces cerevisiae* em 1996 foi um marco importante para o início do sequenciamento de genomas eucarióticos, de maneira geral, e fúngicos, em particular (GOFFEAU et al., 1996). Atualmente, o número de sequências genômicas de fungos nos bancos de dados têm se expandido numa velocidade cada vez maior, especialmente após as iniciativas de empresas privadas e públicas de realizar projetos de sequenciamento desses organismos, como é o caso dos projetos *Broad Institute's Fungal Genome Initiative*, *The 1000 Fungal Genomes Project* e *Génolevures* (GRIGORIEV et al., 2014; ARAÚJO; MAIA, 2018; JAMES, 2020).

Um maior número de genomas de fungos sequenciados significa uma maior possibilidade de entendimento da diversidade molecular em nível de genes, dos processos metabólicos, do modo de vida dos fungos, das adaptações que permitiram a alta variabilidade bioquímica, fisiológica e molecular dentro do reino e de diversas outras questões que o estudo da genômica fúngica permite investigar (MOHANTA; BAE, 2015; NARANJO-ORTIZ; GABALDÓN, 2019).

A estimativa é que existam entre 2,2 e 3,8 milhões de espécies de fungos no planeta, dentre os quais apenas cerca de 100.000 foram descritas. Já a quantidade de genomas completos disponíveis nos bancos de dados não chega nem a 1% das espécies descritas, o que demonstra um grande desafio na utilização das informações disponíveis para o entendimento acerca da genômica fúngica e resolução dos problemas emergentes causados por estes organismos (HAWKSWORTH, 2012; HAWKSWORTH, 2017; WU et al., 2019).

Assim, a partir dos dados genômicos disponíveis no estudo de Stajich (2017), sabe-se que o genoma dos fungos pode ter uma variação de tamanho de até 3 ordens de magnitude, em que o menor genoma amostrado (dentre 325 fungos filogeneticamente diversos) varia de 2 megabases em Microsporidia até 2 gigabases em Pucciniales, com uma mediana de 35 megabases. Já em relação aos genes codificadores de proteínas, se chegou a uma média de 11.000 genes, com uma variação que vai de 1800 genes em Microsporidia a 35.000 genes em

Agaricomycotina (subfiló de Basidiomycota). Esse estudo também mostrou que linhagens que possuem organismos unicelulares (leveduras) tendem a ter genomas menores, uma redução que aparentemente ocorreu de forma independente várias vezes na história destes organismos (JAMES, 2020).

Tomando como referência o tamanho dos genomas das plantas e animais, o tamanho do genoma dos fungos é relativamente pequeno. Entretanto, conforme descrito na seção 1.1 esse fator não diminui a diversidade deste reino (TUNLID; TALBOT; 2002). O estudo de Galagan e colaboradores (2005) demonstrou que diversos membros de um mesmo gênero podem apresentar uma divergência considerável entre seus genomas. Em uma das análises, foram comparados os genomas de *Aspergillus nidulans*, *Aspergillus fumigatus* e *Aspergillus oryzae*, e os resultados mostram que estes organismos compartilham apenas 68% de identidade de sequência de aminoácidos.

Tanto as mudanças genômicas em pequena escala (nível de nucleotídeos / genes) quanto em grande escala (nível de vias bioquímicas / cromossômico) dos organismos fúngicos contribuíram significativamente para sua variação genotípica, fenotípica e para os eventos de especiação. A transferência horizontal de genes entre espécies e a expansão e contração das regiões gênicas também contribuem para a evolução do genoma fúngico. Eventos de alteração cromossômica, duplicação e translocação acontecem com alta frequência especialmente em leveduras, assim como a perda massiva de genes, o que significa que os genomas dos fungos são naturalmente muito dinâmicos (DUNHAM et al., 2002; KOSZUL et al., 2004; MOHANTA; BAE, 2015).

Os fungos, assim como os demais eucariotos, possuem regiões intrônicas e mecanismos de regulação de processamento de RNA mensageiro (mRNA), como o processamento alternativo. Este mecanismo possibilita múltiplas isoformas a partir de um único gene codificador, o que aumenta a diversidade, complexidade e adaptação rápida a novos nichos ou estresses ambientais (KELEMEN et al., 2013; LE et al., 2015; SIEBER et al., 2018). Embora o processamento de mRNA em fungos ainda não tenha sido tão estudado como nos demais eucariotos, já se sabe que este fenômeno tende a ocorrer mais em fungos patogênicos do que em fungos não patogênicos. Assim, devido a presença de regiões gênicas intrônicas, o genoma dos fungos patogênicos geralmente é significativamente maior do que os demais fungos (GRUTZMANN et al., 2014; SIEBER et al., 2018).

1.4 Genômica comparativa

Com o surgimento e aprimoramento dos métodos de sequenciamento genômico, o estudo da natureza e evolução das espécies a partir do sequenciamento completo de seu material genético se torna progressivamente uma alternativa mais viável. Logo, a obtenção dos genomas completos de diferentes organismos tornou possível o estudo comparativo entre as sequências genômicas e identificação de similaridades e particularidades entre os diferentes organismos (YANDELL e ENCE, 2012; WATSON et al., 2015).

Outra vantagem proporcionada por tais tecnologias é a disponibilização de dados genômicos de organismos patogênicos, os quais causam grande impacto na saúde humana, animal e vegetal. O estudo destes dados possibilita uma melhor compreensão dos mecanismos moleculares que contribuem para a infecção por diversos agentes, seja na investigação dos fatores que tornam um organismo patogênico ou inofensivo, seja como determinados genes evoluíram de forma a tornar cepas da mesma espécie mais ou menos virulentas. Assim, a detecção, via genômica comparativa, de genes e vias bioquímicas exclusivas ou enriquecidas em organismos patogênicos fornece uma importante ferramenta para elucidar os mecanismos moleculares da patogenicidade (CHAKRAVORTY et al., 2007; YADAV et al., 2013; ZALUGA et al., 2014). De maneira análoga, o estudo sistemático dos grupos de genes exclusivamente encontrados nos fungos, bem como o seu contextobiológico, fornece importantes mecanismos para o estudo do *modus operandi* desses organismos.

Para realizar estudos da genômica comparativa, é essencial identificar, dentro de um mesmo genoma, bem como entre diferentes espécies, as regiões gênicas de ancestralidade comum, chamadas de regiões homólogas, ou, no caso de regiões codificadoras, genes homólogos codificadores de proteínas. Neste trabalho, buscamos encontrar grupos de homólogos específicos ou super representados nos fungos através da estratégia de genômica comparativa. Assim, as próximas seções descrevem em detalhes conceitos importantes para tal metodologia..

1.5 Conceitos de evolução de genes - homólogos, ortólogos, parálogos e xenólogos

Classicamente, o termo “homologia” foi utilizado para descrever estruturas morfológicas em diferentes espécies que derivam de uma mesma estrutura ancestral encontrada no ancestral comum das espécies em questão, como os membros superiores de tetrápodes. Posteriormente, o conceito de homologia foi estendido aos genes, sendo utilizado para se referir aos genes que foram originados a partir de ancestralidade comum e que podem ocorrer por meio de três fenômenos que produzem as cópias obtidas a partir do gene ancestral comum (Figura 1): especiação (ortólogos), duplicação (parálogos) ou transferência gênica horizontal (xenólogos) (FITCH, 1970; KOONIN, 2005; ALTENHOFF; GLOVER; DESSIMOZ, 2019).

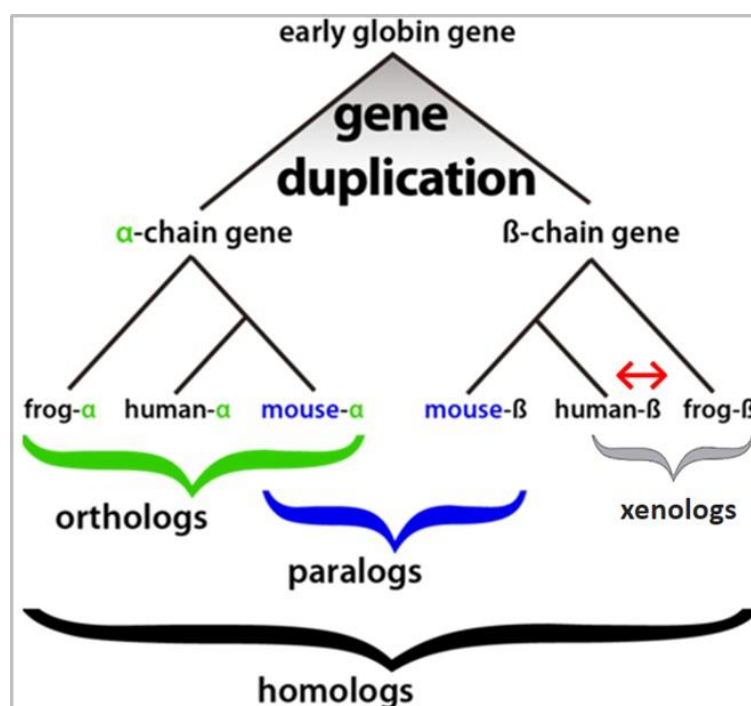


Figura 1. Esquema dos tipos de homologia entre genes da proteína globina
Fonte: Popo H. Liao (modificado)

Desta forma, ortólogos são genes que divergiram por meio da especiação, em que cada descendente possui uma única cópia deste gene. Por compartilharem uma história evolutiva comum, considera-se normalmente que tenham função biológica equivalente nas espécies subjacentes. Assim, genes ortólogos, são ferramentas importantes para os estudos de evolução gênica, análise comparativa entre

genomas e predição de funções desconhecidas de proteínas (KOONIN, 2005; MAZZA et al., 2009; ALTENHOFF; GLOVER; DESSIMOZ, 2019).

Já os parálogos são os genes que divergiram por meio de eventos de duplicação, podendo então ser encontradas mais de uma cópia no mesmo genoma. Este tipo de divergência comumente tende a levar à realocação de funções (através de mutações ou recombinações) e formação de famílias gênicas, algumas com centenas de membros (KOONIN, 2005; MAZZA et al., 2009; ALTENHOFF; GLOVER; DESSIMOZ, 2019).

Por fim, temos os genes xenólogos, que se referem aos genes que foram adquiridos de outras espécies por meio de transferência horizontal, também conhecida por transferência lateral. Podemos considerar xenologia se a origem do ancestral comum de pelo menos um do par de genes contém transferência de material genético interespecie. Esta classe de homologia é a única em que o gene é descendido de um organismo para o outro, e, o reconhecimento de genes xenólogos é crucial para se evitar filogenias equivocadas (FITCH, 2000; WEBBER; PONTING, 2004).

1.6 Bancos de dados de vias bioquímicas - KEGG

O KEGG (Enciclopédia Kyoto de Genes e Genomas) é um repositório composto por vários bancos de dados contendo informações padronizadas e conectadas entre si. Estas informações estão disponíveis a fim de serem usadas em análises dos diversos e crescentes dados gerados a partir dos avanços das tecnologias de sequenciamento de genoma. A proposta do KEGG é disponibilizar dados biológicos e dados de funções moleculares de alto nível através do recurso computacional, por meio do cruzamento das informações de genes, reações e interações moleculares, vias metabólicas, bem como funcionalidades envolvidas (KANEHISA; GOTO, 2000; KANEHISA et al., 2016; 2021).

Para isso, o sistema possui uma arquitetura que integra e organiza 18 bancos de dados (Figuras 2) de forma sistemática por meio de quatro categorias: Sistemas de Informação (*KEGG Pathway, Brite e Module*); Informações Genômicas (*KEGG Orthology - KO, Genome, Genes e SSDB - Sequence Similarity DataBase*); Informações Químicas – *KEGG LIGAND (Compound, Glycan, Reaction, RClass, Enzyme)*; e Informações de Saúde (*KEGG Network, Variant, Disease, Drug, Dgroup*,

Environ, Japic e DailyMed) (KANEHISA; GOTO, 2000; KANEHISA et al., 2016; 2021).

Category	Database name	Content	Prefix
Systems Information	KEGG PATHWAY	KEGG pathway maps	map, hsa, etc.
	KEGG BRITE	BRITE functional hierarchies and tables	br, ko, etc.
	KEGG MODULE	KEGG modules Reaction modules	M RM
Genomic Information	KEGG ORTHOLOGY (KO)	KO groups for functional orthologs	K
	KEGG GENOME	KEGG organisms and viruses	T
	KEGG GENES	Genes and proteins	
	KEGG SSDB	Sequence similarity among GENES entries	
Chemical Information	KEGG COMPOUND	Metabolites and other small molecules	C
	KEGG GLYCAN	Glycans	G
	KEGG REACTION	Biochemical reactions	R
	KEGG RCLASS	Reaction class	RC
	KEGG ENZYME	Enzyme nomenclature	
Health Information	KEGG NETWORK	Disease-related network elements	N
		Network variation maps	nt
	KEGG VARIANT	Human gene variants	
	KEGG DISEASE	Human diseases	H
	KEGG DRUG	Drugs	D
	KEGG DGROU	Drug groups	DG
	KEGG ENVIRON	Crude drugs and health-related substances	E
	JAPIC	Japanese drug labels	
	DailyMed	FDA drug labels (links only)	

Figura 2. Índice do Bancos de Dados KEGG - O KEGG é dividido em 4 grandes categorias, destacadas em verde (Sistemas de Informação), vermelho (Informação Genômica), azul claro (Informação Química) e roxo (Informação de Saúde), que possuem diferentes tipos de entrada, conteúdo e prefixos para cada categoria.

Fonte: KANEHISA, 2021.

Cada tipo de entrada recebe um identificador KEGG composto por um prefixo do banco de dados (Figura 2) ao qual pertence acrescido de seu número KEGG. Por exemplo, a via bioquímica da Glicólise/Gliconeogênese tem como identificador “map00010”, em que “map” é o prefixo do banco de dados e “00010” é sua numeração registrada no KEGG. Em nosso trabalho, utilizamos dois bancos de dados KEGG que são explicados em detalhes nas próximas seções: KEGG pathway KEGG orthology.

1.6.1 KEGG Pathway

KEGG Pathway é um conjunto de mapas de vias metabólicas, desenhadas via curadoria manual, que representam as interações moleculares, reações e redes de relações que envolvem o metabolismo, processamento genético, ambiental, entre outros. Cada mapa de via é representado pelo identificador em que contém o prefixo

de 2 a 4 letras mais um número de cinco dígitos (KANEHISA; GOTO, 2000. KANEHISA, 2021).

O KEGG Pathway pode ser considerado como o núcleo do banco de dados KEGG, uma vez que, através de um conjunto de mapas de vias, integra todos os demais dados das quatro categorias do KEGG. Dessa forma, este banco de dados possui em seu conteúdo informações sobre vias metabólicas e vias regulatórias, que por sua vez possuem dados de processamento de informações genéticas, ambientais, processos celulares e doenças humanas. o KEGG Pathway também possui informações sobre as relações estruturais das vias por meio de um mapa da estrutura de drogas. Assim, cada um dos componentes das vias, no que lhe concerne, consiste em uma entrada (link) que nos direciona para uma nova página do KEGG, a qual contém informações específicas acerca desse componente molecular, como é exemplificado na Figura 4 (OGATA et al., 1998; AOKI-KINOSHITA; KANEHISA, 2007; KANEHISA, 2020).

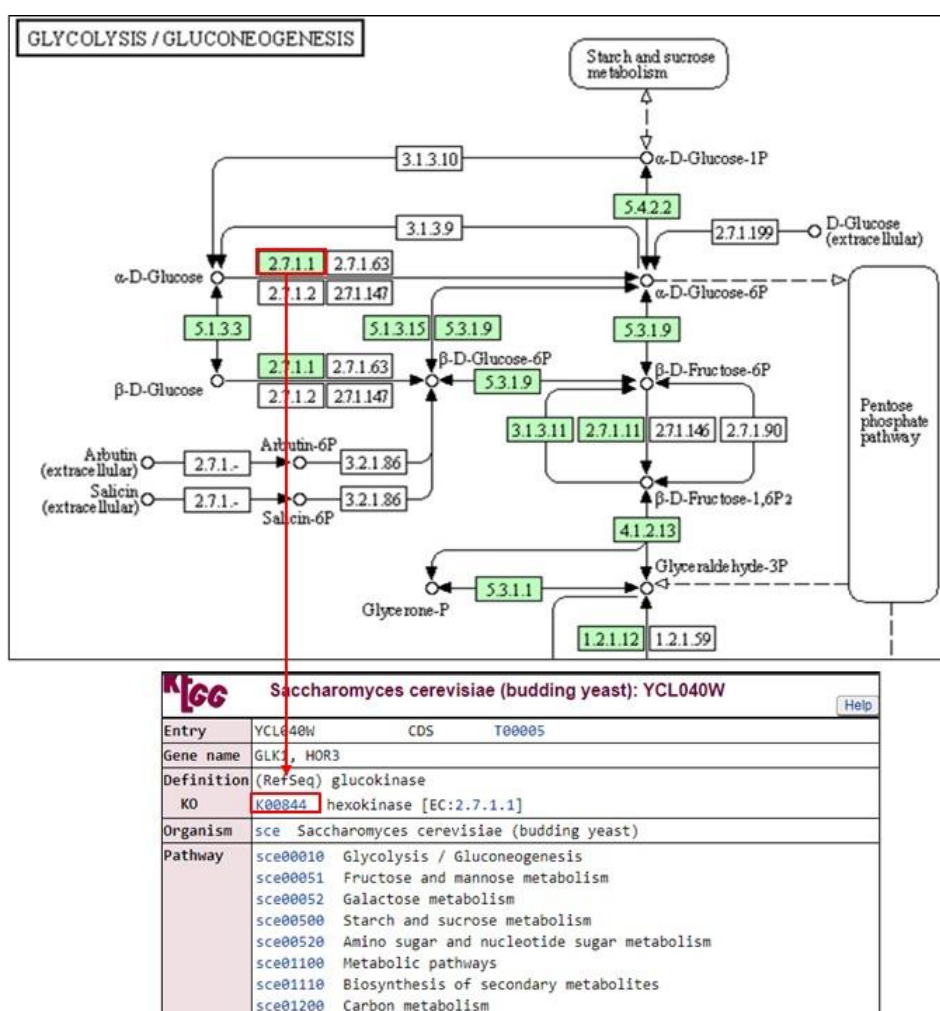


Figura 3. Representação KEGG da via da glicólise e gliconeogênese e página de informações - Os mapas das vias do KEGG são representados por elementos gráficos que simbolizam os metabólitos (círculos pequenos), as reações enzimáticas (retângulos), o sentido em que acontecem as reações (setas) e as vias metabólicas associadas (caixas maiores). Em verde estão destacadas todas as reações enzimáticas que ocorrem em *Saccharomyces cerevisiae*. Ao se clicar em um retângulo, automaticamente somos direcionados para uma nova página com informações detalhadas acerca de tal enzima.

Fonte: KEGG Pathway

Na figura 3, podemos observar as reações que compõem a via da glicólise e gliconeogênese, que possui o identificador map00010. Nas vias metabólicas KEGG, os pequenos círculos representam os metabólitos, as setas representam o sentido da reação, os retângulos menores representam as reações enzimáticas e os maiores representam as vias relacionadas. Ao clicarmos nas figuras dos compostos somos direcionados para a página que descreve tal composto. No exemplo em questão, estão destacados em verde as enzimas presentes em *Saccharomyces cerevisiae* na via da glicólise/gliconeogênese, e logo abaixo uma página de redirecionamento do KEGG que traz as informações detalhadas da enzima 2.7.1.1 (hexokinase).

1.6.2 KEGG Orthology

Para integrar a rede de vias metabólicas KEGG, este banco de dados utiliza informações de homologia, representadas através de grupos de genes homólogos funcionais, o qual é denominado KEGG Orthology (KO). Apesar do nome, o banco integra dados de genes homólogos, e não somente ortólogos, os quais são definidos manualmente através do contexto das redes de vias KEGG (KEGG Pathway). Cada retângulo na via (Figura 4) representa uma reação enzimática que recebe um identificador, o qual pode corresponder a um ou mais grupos de genes homólogos. Esse identificador (também chamado de número K) é constituído pela letra k acrescida cinco números, simbolizando o homólogo funcional que codifica a reação daquela etapa da via metabólica (KANEHISA et al., 2016).



Figura 4. **Esquema da integração dos bancos de dados KEGG por meio do KEGG Orthology**
 Fonte: KANEHISA, 2020.

Diversos dados de interesse para genômica comparativa podem ser obtidos dentro do banco de dados KEGG Genes, que contém conjuntos de genes, bem como suas respectivas anotações, as quais possuem correspondências dentro das vias KEGG de acordo com sua homologia. Desse modo, a partir de evidências científicas, os mapas KEGG são projetados de forma a fornecer informações de compartilhamento de vias entre diferentes organismos, que correspondem aos genes homólogos. Estes grupos de genes que possuem homologia entre espécies são definidos como Kegg Orthology (KO), e cada grupo KO traz consigo informações das vias e das funções enzimáticas correlacionadas (KANEHISA et al., 2014).

A caracterização dos KEGG Orthology é feita a partir de evidências experimentais de sequências de genes e proteínas funcionalmente caracterizadas em organismos específicos, que são então usados para atribuir genes ortólogos em outros organismos com base na similaridade de sequência, bem como genes parálogos no mesmo organismo. A anotação dos genomas é feita na forma de atribuição de identificadores a genes individuais no genoma, em substituição a uma descrição textual (KANEHISA et al., 2016; KANEHISA; SATO; MORISHIMA, 2016).

1.6.3 KEGG API e Protocolo REST

O banco de dados KEGG disponibiliza uma interface de programação de aplicativos (*Application Programming Interface*, API) do tipo REST (*Representational State Transfer*) chamado KEGG API. Arquiteturas de software do tipo REST compreendem um protocolo sistematizado para a troca de informações entre computadores em serviços web, (FIELDING, 2000). Uma interface de programação

de aplicativos (API) é uma interface de computação que permite a troca de informações entre computadores de forma definida. O KEGG utiliza a API como forma de possibilitar aos usuários o acesso aos recursos armazenados no banco de dados de maneira programática, de forma que o usuário possa acessar recursos disponíveis no banco de dados KEGG em seus próprios programas internos, e para investigar seus próprios dados. Cada vez que a KEGG API é utilizada, os dados acessados são os mesmos dos atualizados no servidor da plataforma (KAWASHIMA; KATAYAMA, 2003).

1.7 Reações de *chokepoints*

Para que um patógeno se desenvolva, se multiplique e cause infecção, é necessário que todo seu potencial metabólico seja alcançado. Dessa forma, toda a rede de reações enzimáticas que fazem parte das vias metabólicas essenciais para o parasita são cruciais para a viabilidade do mesmo e são, portanto, interessantes alvos de drogas (BAYLY; MACREADIE, 2002; UMEDA et al., 2011; TAYLOR et al., 2013). O conhecimento gerado a partir dos estudos genômicos e bioquímicos que envolvem organismos patogênicos possibilita que novos alvos de drogas sejam descobertos, e que se amplie as frentes de combate a doenças infecciosas em animais, plantas e seres humanos (TYAGI et al., 2019).

As reações enzimáticas cruciais para a viabilidade dos organismos foram estudadas por Yeh et al. (2004) e então denominadas como “*chokepoints*” (Figura 5). Reações do tipo *chokepoints* são definidas por ou consumirem um substrato exclusivo (Figura 5-A) ou produzirem um produto exclusivo (Figura 5-B) as quais, ao serem inibidas, promovem a inibição de toda a cascata de reações posteriores. A falta de caminhos alternativos para que uma reação aconteça é o que faz de uma enzima um *chokepoint*. A premissa desta ideia é de que, se a enzima responsável por essa reação for inibida, então toda a via bioquímica será inviabilizada, levando ao acúmulo de substrato ou deixando de obter o produto que realizaria a próxima etapa da via (RAHMAN, 2006; PALUMBO, 2007; TAYLOR et al., 2013).

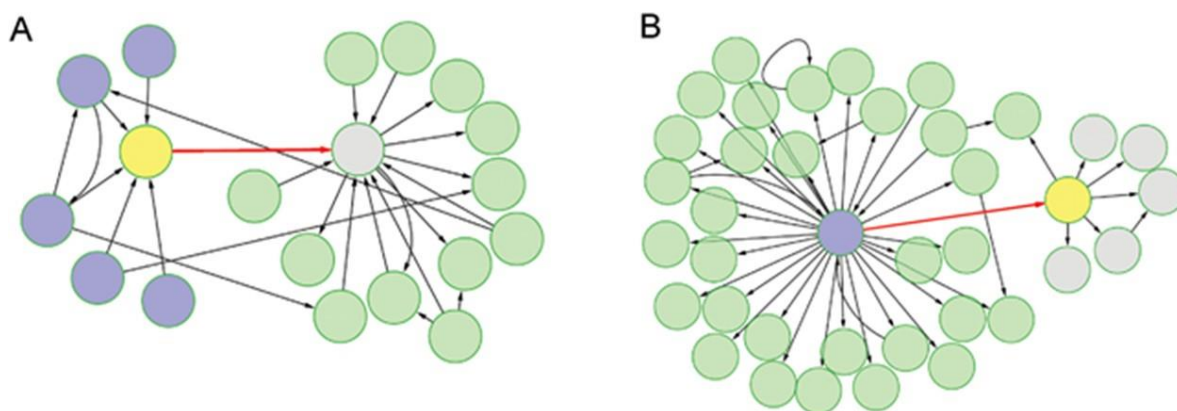


Figura 5. Demonstração gráfica dos tipos de reações de *chokepoints* - Para que uma reação seja um *chokepoint*, a mesma deve ou consumir um substrato exclusivo (A) ou produzir um produto exclusivo (B). As bolinhas amarelas representam os *chokepoints* e as setinhas vermelhas representam a reação que deixa de acontecer quando o *chokepoint* é inibido.
Fonte: Taylor et al., 2013 (modificado).

Chokepoints são candidatos interessantes a alvos de medicamentos devido ao seu potencial de prejudicar seletivamente os patógenos parasitas sem prejudicar o hospedeiro. Para isso, o *chokepoint* em questão deve acontecer exclusivamente no organismo parasita e então, ao ser inibida, o hospedeiro não sofre com os efeitos adversos ou toxicidade do inibidor (TYAGI et al., 2019). Atualmente já existem drogas voltadas para *chokepoints* no combate a doenças de grande impacto na saúde pública, como malária e tuberculose (BAYLY; MACREADIE, 2002; TIMMINS; DERETIC, 2006; UMEDA et al., 2011).

2 OBJETIVOS

2.1 Objetivo geral

Obter a assinatura de genes, vias e processos metabólicos enriquecidos nos genomas dos fungos quando comparados aos demais eucariotos (genômica comparativa funcional).

2.2 Objetivos específicos

- Obter os genomas completos de eucariotos KEGG
- Avaliar os KOs super-representados nos fungos em comparação aos eucariotos não fungos;
- Mapear os KOs dos genomas dos fungos nas vias KEGG;
- Identificar os KOs significativamente enriquecidos em fungos em relação aos demais eucariotos;
- Buscar por candidatos enzimáticos enriquecidos nos fungos para propor *chokepoints*.

3 MÉTODOS

Os métodos utilizados neste projeto foram organizados em sete etapas ao longo do seguinte workflow:

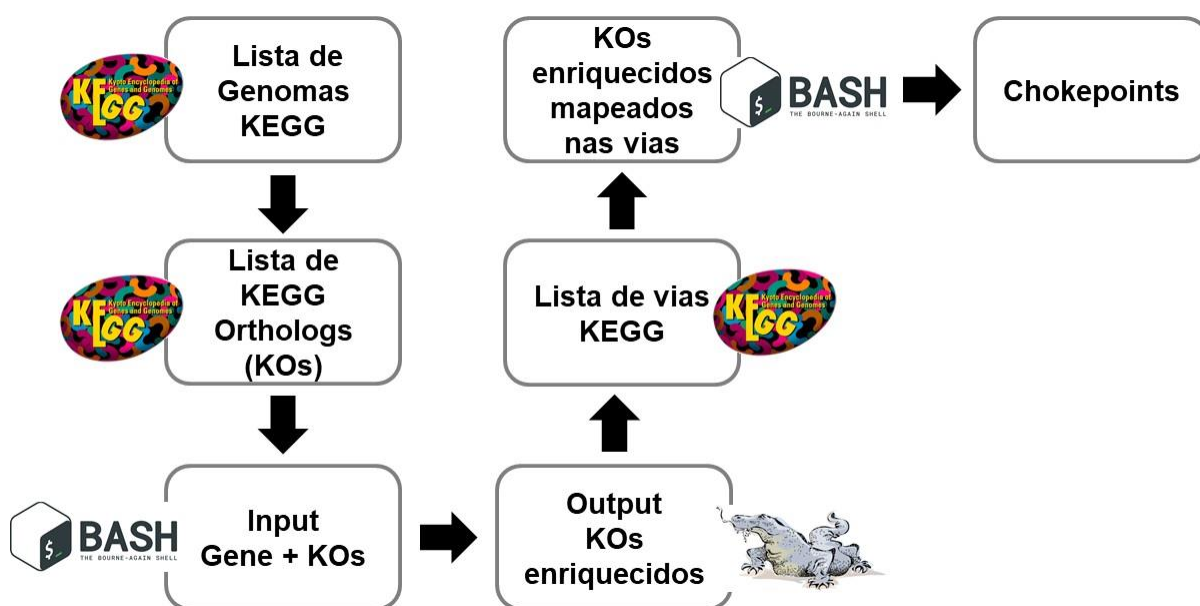


Figura 6. Workflow das etapas da obtenção dos dados deste projeto
 Fonte: Dados da pesquisa

3.1 Estrutura computacional utilizada neste projeto

As operações deste projeto foram desenvolvidas através de um servidor DELL composto por 2 processadores Intel Xeon E5-4610 v2 2.3GHz, totalizando 64 threads. Este equipamento conta também com 128GB de memória RAM e com um sistema operacional CentOS Linux, versão 7.5.1804 com suporte às linguagens de programação PERL5 versão 16, subversão 3 (v5.16.3) e R versão 3.5.0. O programa CALANGO, que utilizamos para as análises dos dados, se encontra instalado neste servidor.

3.2 Obtenção dos organismos presentes no KEGG e utilizados nesse estudo

Os genomas estudados foram obtidos através do banco de dados KEGG em outubro de 2019. Para realizarmos a extração dos dados genômicos, utilizamos o KEGG API, que compreende uma interface de programação para permitir ao usuário interagir com o banco KEGG para realizar o download, manipulação e integração de

dados deste banco através de um computador com internet. Os arquivos da plataforma são do tipo *REST*, um tipo de arquitetura de software que permite a comunicação entre computadores via rede.

A partir do protocolo *REST*, fizemos o download dos arquivos dos genomas de todos os organismos disponíveis na plataforma em 10/10/2019. A contagem do arquivo nos retornou 6330 linhas, logo, temos essa quantidade de organismos no banco de dados KEGG, dentre procariotos e eucariotos disponíveis na plataforma.

wget http://rest.kegg.jp/list/organism

T01358	ota	Ostreococcus tauri	Eukaryotes;Plants;Green algae
T03089	bpg	Bathycoccus prasinos	Eukaryotes;Plants;Green algae
T02695	mis	Micromonas commoda	Eukaryotes;Plants;Green algae
T02696	mpp	Micromonas pusilla	Eukaryotes;Plants;Green algae
T00175	cme	Cyanidioschyzon merolae	Eukaryotes;Plants;Red algae
T02926	gsl	Galdieria sulphuraria	Eukaryotes;Plants;Red algae
T02916	ccp	Chondrus crispus (carragheen)	Eukaryotes;Plants;Red algae
T00005	sce	Saccharomyces cerevisiae (budding yeast)	Eukaryotes;Fungi;Ascomycetes
T00167	ago	Ashbya gossypii (Eremothecium gossypii)	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T02237	erc	Eremothecium cymbalariae	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01025	kla	Kluyveromyces lactis	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T05269	kmx	Kluyveromyces marxianus	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01117	lth	Lachancea thermotolerans	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01049	vpo	Vanderwaltozyma polyspora	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01118	zro	Zygosaccharomyces rouxii	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01024	cgr	Candida glabrata	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T02233	ncs	Naumovozyma castelli	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T02234	ndi	Naumovozyma dairenensis	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T02235	tpf	Tetrapisispora phaffii	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T02489	tbl	Tetrapisispora blattae	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T02236	tdl	Torulaspora delbrueckii	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T02487	kaf	Kazachstania africana	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01092	ppa	Komagataella phaffii	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01026	dha	Debaryomyces hansenii	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01023	pic	Scheffersomyces stipitis	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01150	pgu	Meyerozyma guilliermondii	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T03101	spaa	Spathaspora passalidarum	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01116	lel	Lodderomyces elongisporus	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T00189	cal	Candida albicans	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes
T01115	ctp	Candida tropicalis	Eukaryotes;Fungi;Ascomycetes;Saccharomycetes

Figura 7. Arquivo dos genomas contendo o ID, código e descrição de cada organismo contido no KEGG - O arquivo de genomas possui dados tabulares em que a primeira coluna representa a lista de entradas (ID) KEGG de cada organismo, a segunda representa o código de cada organismo, a terceira contém o nome científico e, a quarta, os grupos filogenéticos aos quais cada um pertence. Fonte: Dados da pesquisa

3.3 Seleção de genomas de fungos e demais eucariotos disponíveis no KEGG

De posse de todos os genomas celulares disponíveis no KEGG, procedemos realizando a seleção dos organismos eucariotos e, a partir desses, criamos 2 novos arquivos do tipo lista contendo, respectivamente, os seguintes subconjuntos de eucariotos: Fungos e Não-Fungos. O arquivo de saída continha 535 linhas, significando ser esta a quantidade de genomas de Eucariotos.

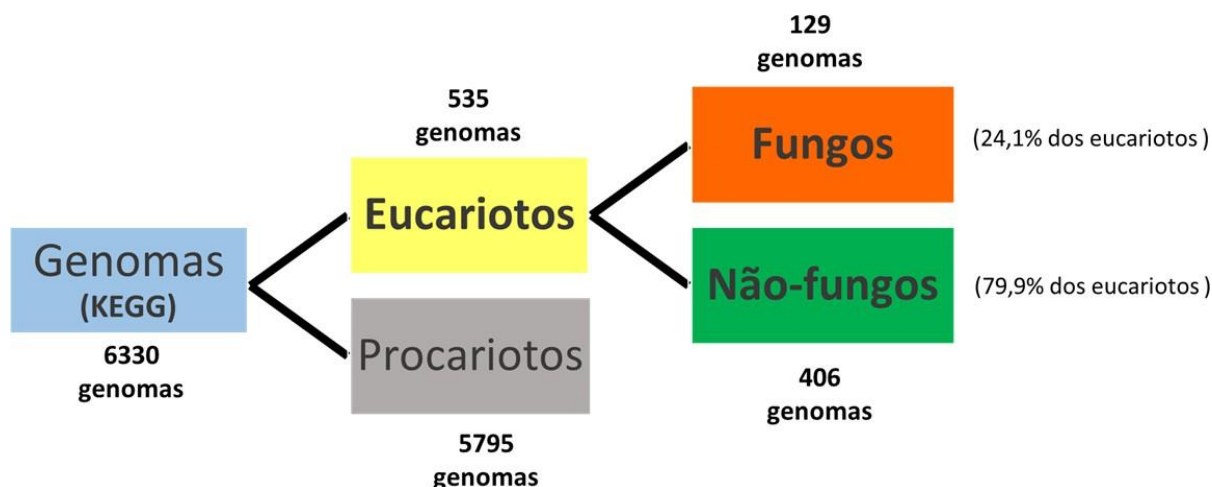


Figura 8. Etapas da obtenção dos genomas dos organismos utilizados neste estudo - Os genomas eucarióticos utilizados nesta pesquisa foram extraídos a partir de um arquivo contendo todos os genomas do KEGG (em azul). O arquivo contendo apenas genomas de eucariotos foi subdividido em dois grupos: Fungos (em laranja) e Não-fungos (em verde)
Fonte: Dados da pesquisa

Para preparar os arquivos contendo os genomas de Fungos e não-Fungos, ambos necessários como entrada para o programa CALANGO, realizamos a separação dos dados genômicos de eucariotos em 2 grupos: fungos e não fungos. O arquivo de saída do primeiro grupo continha 129 espécies. Já o total de espécies do arquivo de saída do segundo grupo foi de 406, que somados aos fungos totalizaram os 535 organismos baixados inicialmente. Os comandos linux utilizados para obter as listas de organismos eucarióticos fungos e não-fungos seguem abaixo:

```
grep 'Fungi;' organism > Fungi_info_list.txt
```

```
grep 'Eukaryotes;' organism | grep -v 'Fungi;' > Eukanofungi_info_list.txt
```

A partir dos novos arquivos, criamos duas novas listas contendo apenas a primeira coluna de cada arquivo original, que contém os identificadores (IDs) únicos de cada um dos organismos que constam nos arquivos dos fungos e dos não fungos. Estas duas novas listas chamamos de: ID_Fungi.txt e ID_EukanoFungi.txt.

```
cat Eukanofungi_info_list.txt | awk '{print$1}' > ID_Eukanofungi.txt
```

```
cat Fungi_info_list.txt | awk '{print$1}' > ID_Fungi.txt
```

3.4 Obtenção dos grupos de homólogos (KOs) dos organismos

A partir de cada um dos IDs de espécies obtidos na etapa anterior, procedemos com a obtenção de todos os grupos de genes homólogos presentes nestes genomas, que no KEGG são anotados utilizando identificadores KO (KEGG Orthology). Resumidamente, para cada um desses organismos, existe um determinado número de KOs associados, os quais podem estar relacionados a um ou mais genes no mesmo indivíduo, e podem ser encontrados em uma ou em várias espécies simultaneamente. Assim, fizemos um *script* para, via KEGG API, realizar o download dos arquivos KO para cada organismo a partir das nossas listas de IDs.

```
for i in `less ID_Fungi.txt`; do echo wget http://rest.kegg.jp/link/ko/$i ; done >
download_Fungi.sh
```

```
for i in `less ID_Eukanofungi.txt`; do echo wget http://rest.kegg.jp/link/ko/$i ;
done > download_Eukanofungi.sh
```

O código acima gerou 2 arquivos “.sh”, os quais tornamos executáveis, de modo a automatizar o processo de obtenção dos KOs para cada genoma. Procedemos com a criação de dois novos diretórios para armazenar os KOs dos dois grupos de organismos. Movemos os arquivos “.sh” em cada um desses diretórios. Desta forma, ao serem executados, estes arquivos “.sh” realizaram o download dos registros de KOs de cada organismo (ID) nos diretórios em questão.

Os arquivos gerados para cada organismo compreendem uma lista contendo todos os IDs dos genes de cada grupo de eucariontes, bem como seus respectivos KOs associados. Dentro de cada arquivo da lista foi gerada uma tabela na qual a primeira coluna corresponde aos genes do organismo (geneID) e a segunda coluna representa os Kegg Orthology (KO) presentes em tais genes. Os diretórios contendo os arquivos (ID+KOs) foram chamados de Fungi_ID_KO e Eukanofungi_ID_KO.

Dos 129 genomas de Fungos, realizamos o download satisfatório de 128. Ao comparar os arquivos, detectamos que o organismo faltante é o fungo filamentoso *Scedosporium apiospermum* (T06367), que foi descartado do estudo, uma vez que a interface KEGG API não foi capaz de obter as informações de seus genes e KOs.

Ao final dessa etapa, obtivemos, para um total de 120 genomas de fungos e 406 genomas de eucariotos não-fungos, um arquivo contendo os identificadores dos genes em cada organismo, bem como a anotação dos grupos KOs aos quais os mesmos fazem parte.

3.5 Curadoria dos arquivos (Identificadores + KOs)

Após a etapa anterior, realizamos a curadoria das tabelas geradas até o momento. Primeiramente, para cada arquivo representando os genes de cada espécie, substituímos a primeira coluna destes arquivos (identificadores dos genes) por uma sequência numérica, para que seja possível executar nossas análises pelo software CALANGO. Para isso, escrevemos um programa chamado “*fix.pl*” utilizando a linguagem de programação PERL, e cujo código segue abaixo.

```
use strict;
use warnings;

open(IN, $ARGV[0]);

my $i = 0;

my $header = <IN>;

print $header;

while (my $line = <IN>) {
    chomp $line;
    my @aux = split(/\t/, $line);
    my $KO = $aux[1];
    print("$i\t$KO\n");
    $i++;
}

close IN;
```

Em seguida, retiramos os caracteres “ko:” da segunda coluna destes arquivos, a qual contém os identificadores KO, além de inserimos um cabeçalho que identifica o conteúdo de cada coluna. Ao ser criado, o arquivo de cabeçalho foi copiado para os diretórios de cada grupo de organismos, e dentro de cada um dos

diretórios, utilizamos um comando para concatenar o arquivo “header” a cada tabela contida nos arquivos dos IDs, de modo a gerar os novos arquivos temporários (com o acréscimo “.comheader” ao nome para diferenciar dos arquivos originais), que foram retirados do diretório e armazenados num diretório de dados temporários. O código para a realização destas operações segue abaixo:

```
for `i in T*`; do sed -i -e 's/ko://g' ${i} ;done
```

```
for `i in `ls T*`; do cat header $i > $i.comheader ; done
```

bor:COCMIDRAFT_4610	ko:K24965	geneID	K0
bor:COCMIDRAFT_5247	ko:K06116	bor:COCMIDRAFT_5247	K06116
bor:COCMIDRAFT_32608	ko:K11567	bor:COCMIDRAFT_32608	K11567
bor:COCMIDRAFT_83419	ko:K11109	bor:COCMIDRAFT_83419	K11109
bor:COCMIDRAFT_81043	ko:K24104	bor:COCMIDRAFT_7775	K19985
bor:COCMIDRAFT_7775	ko:K19985	bor:COCMIDRAFT_1877	K01673
bor:COCMIDRAFT_1877	ko:K01673	bor:COCMIDRAFT_81564	K00915
bor:COCMIDRAFT_81564	ko:K00915	bor:COCMIDRAFT_101481	K22063
bor:COCMIDRAFT_101481	ko:K22063	bor:COCMIDRAFT_91647	K22746
bor:COCMIDRAFT_4145	ko:K24348	bor:COCMIDRAFT_101286	K22484
bor:COCMIDRAFT_91647	ko:K22746	bor:COCMIDRAFT_37874	K06910
bor:COCMIDRAFT_101286	ko:K22484	bor:COCMIDRAFT_95822	K19562
bor:COCMIDRAFT_79677	ko:K14818	bor:COCMIDRAFT_4583	K18802
bor:COCMIDRAFT_37874	ko:K06910	bor:COCMIDRAFT_32756	K18802
bor:COCMIDRAFT_95822	ko:K19562	bor:COCMIDRAFT_27414	K07513
bor:COCMIDRAFT_4583	ko:K18802	bor:COCMIDRAFT_4198	K01536
bor:COCMIDRAFT_32756	ko:K18802	bor:COCMIDRAFT_108172	K00249
bor:COCMIDRAFT_27414	ko:K07513	bor:COCMIDRAFT_43615	K03880
bor:COCMIDRAFT_4198	ko:K01536	bor:COCMIDRAFT_90171	K07975
bor:COCMIDRAFT_108172	ko:K00249	bor:COCMIDRAFT_105690	K07975
bor:COCMIDRAFT_43615	ko:K03880	bor:COCMIDRAFT_3064	K09202
bor:COCMIDRAFT_90171	ko:K07975	bor:COCMIDRAFT_90841	K01210
bor:COCMIDRAFT_105690	ko:K07975	bor:COCMIDRAFT_5182	K01426
bor:COCMIDRAFT_3064	ko:K09202	bor:COCMIDRAFT_92457	K19564
bor:COCMIDRAFT_90841	ko:K01210	bor:COCMIDRAFT_29151	K22315
bor:COCMIDRAFT_5182	ko:K01426	bor:COCMIDRAFT_98164	K18932
bor:COCMIDRAFT_92457	ko:K19564	bor:COCMIDRAFT_4932	K18664
bor:COCMIDRAFT_29151	ko:K22315	bor:COCMIDRAFT_87557	K18932
bor:COCMIDRAFT_98164	ko:K18932	bor:COCMIDRAFT_3506	K10747
bor:COCMIDRAFT_1102	ko:K15435	bor:COCMIDRAFT_85778	K21456
bor:COCMIDRAFT_34515	ko:K24104	bor:COCMIDRAFT_22848	K21452
bor:COCMIDRAFT_4932	ko:K18664	bor:COCMIDRAFT_101174	K07766
bor:COCMIDRAFT_87557	ko:K18932	bor:COCMIDRAFT_109017	K19307
bor:COCMIDRAFT_3506	ko:K10747	bor:COCMIDRAFT_97381	K19308
bor:COCMIDRAFT_85778	ko:K21456	bor:COCMIDRAFT_21488	K19030
bor:COCMIDRAFT_22848	ko:K21452	bor:COCMIDRAFT_98525	K23802
bor:COCMIDRAFT_101174	ko:K07766	bor:COCMIDRAFT_90932	K19953
T03125		T03125.comheader	

Figura 9. Exemplo de um arquivo antes e depois da curadoria

Fonte: Dados da pesquisa

Estando de posse dos arquivos com cabeçalho e substituição da primeira coluna por sequência alfanumérica, estes já estão prontos para serem lidos pelo programa CALANGO. O próximo passo foi preparar o dicionário dos KOs, os quais contém a definição de cada um dos grupos de genes homólogos de acordo com o

banco de dados KEGG, e que também são necessários para a execução de CALANGO.

3.6 Montagem do dicionário de KOs

Para atribuir a descrição de cada Kegg Orthology ao seu identificador, inicialmente criamos uma lista contendo todos os identificadores KO de todos os organismos, de modo que cada identificador ocorra uma única vez nessa lista. Para tal, selecionamos e concatenamos a segunda coluna dos arquivos `ko_Eukanofungi` e `ko_Fungi`. Organizamos a lista em ordem numérica e retiramos as redundâncias. Em seguida, abrimos este arquivo criado e demos o comando de download da descrição de cada KO, de forma que para cada linha a primeira coluna constasse o identificador Kegg Orthology e a segunda coluna fosse composta pela descrição do mesmo.

```
cat * | awk '{print $2}' | sort | uniq > ko_Eukanofungi
```

```
cat * | awk '{print $2}' | sort | uniq > ko_Fungi
```

```
cat ko_Eukanofungi ko_Fungi | sort | uniq > ko_sem_descricao.txt
```

```
wget http://rest.kegg.jp/list/ko
```

O arquivo “ko”, contendo a descrição de todos os KOs e a lista dos KO IDs, foi renomeado para “dicionario_eukaryotes”, completando assim o nosso dicionário. Vale ressaltar que 54 KOs ainda não possuem descrição disponível no KEGG, mas não foram descartados do estudo.

```
grep -f ko_semdescricao ko_organisms > dicionario_eukaryotes
```

```

K00045 E1.1.1.67, mtLK; mannitol 2-dehydrogenase [EC:1.1.1.67]
K00049 GRHPR; glyoxylate/hydroxypyruvate reductase [EC:1.1.1.79 1.1.1.81]
K00051 E1.1.1.82; malate dehydrogenase (NADP+) [EC:1.1.1.82]
K00052 lcuB, IMDH; 3-isopropylmalate dehydrogenase [EC:1.1.1.85]
K00053 ilvC; ketol-acid reductoisomerase [EC:1.1.1.86]
K00058 serA, PHGDH; D-3-phosphoglycerate dehydrogenase / 2-oxoglutarate reductase [EC:1.1.1.95 1.1.1.399]
K00059 fabG, OAR1; 3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]
K00060 tdh; threonine 3-dehydrogenase [EC:1.1.1.103]
K00061 RDH5; 11-cis-retinol dehydrogenase [EC:1.1.1.315]
K00062 E1.1.1.116; D-arabinose 1-dehydrogenase [EC:1.1.1.116]
K00063 ARA1; D-arabinose 1-dehydrogenase [EC:1.1.1.117]
K00064 E1.1.1.122; D-threo-aldose 1-dehydrogenase [EC:1.1.1.122]
K00065 kduD; 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase [EC:1.1.1.127]
K00067 rfbD, rmlD; dTDP-4-dehydrorhamnose reductase [EC:1.1.1.133]
K00069 HPGD; 15-hydroxyprostaglandin dehydrogenase (NAD) [EC:1.1.1.141]
K00070 HSD3B; 3beta-hydroxy-Delta5-steroid dehydrogenase / steroid Delta-isomerase [EC:1.1.1.145 5.3.3.1]
K00071 HSD11B2; corticosteroid 11-beta-dehydrogenase isozyme 2 [EC:1.1.1.-]
K00072 SPR; sepiapterin reductase [EC:1.1.1.153]
K00074 paaH, hbd, fadB, mmgB; 3-hydroxybutyryl-CoA dehydrogenase [EC:1.1.1.157]
K00075 murB; UDP-N-acetylmuramate dehydrogenase [EC:1.3.1.98]
K00077 panE, apbA; 2-dehydropantoate 2-reductase [EC:1.1.1.169]
K00078 DHDH; dihydrodiol dehydrogenase / D-xylose 1-dehydrogenase (NADP) [EC:1.3.1.20 1.1.1.179]
K00079 CBR1; carbonyl reductase 1 [EC:1.1.1.184 1.1.1.189 1.1.1.197]
K00081 CBR2; carbonyl reductase 2 [EC:1.1.1.184]
K00083 CAD; cinnamyl-alcohol dehydrogenase [EC:1.1.1.195]
K00084 CBR3; carbonyl reductase 3 [EC:1.1.1.184]
K00085 E1.1.1.200; aldose-6-phosphate reductase (NADPH2) [EC:1.1.1.200]
K00088 IMPDH, guaB; IMP dehydrogenase [EC:1.1.1.205]
K00089 AKR1C2; 3alpha-hydroxysteroid 3-dehydrogenase [EC:1.1.1.213 1.1.1.357]
K00090 ghrB; glyoxylate/hydroxypyruvate/2-ketogluconate reductase [EC:1.1.1.79 1.1.1.81 1.1.1.215]
K00099 dxr; 1-deoxy-D-xylulose-5-phosphate reductoisomerase [EC:1.1.1.267]
K00101 lldD; L-lactate dehydrogenase (cytochrome) [EC:1.1.2.3]
K00102 LDHD, dld; D-lactate dehydrogenase (cytochrome) [EC:1.1.2.4]
K00103 GULO; L-gulonolactone oxidase [EC:1.1.3.8]
K00106 XDH; xanthine dehydrogenase/oxidase [EC:1.1.1.4 1.1.1.17 1.1.1.3.2]
K00107 ALO; D-arabinono-1,4-lactone oxidase [EC:1.1.3.37]
K00108 beta, CHDH; choline dehydrogenase [EC:1.1.99.1]
:

```

Figura 10. Arquivo contendo os identificadores dos KOs dos eucariotos e suas respectivas descrições - Na primeira coluna consta a lista das entradas KEGG Orthology e na segunda a descrição de cada um deles.

Fonte: Dados da pesquisa

3.7 Classificação dos fungos de acordo com a patogenicidade

A fim de realizarmos nossas análises comparativas, buscamos na literatura dados que comprovem a presença ou ausência de patogenicidade das espécies de fungos utilizadas em nosso estudo. Para isso, utilizamos os IDs de todos os fungos baixados via KEGG juntamente com as informações referentes a cada um, em que constam dentre outras informações o nome científico, que foi a palavra chave utilizada nas buscas por artigos científicos em todos os bancos de dados do Google Scholar.

Dos 129 fungos utilizados em nosso estudo, 72 estão relatados na literatura como patogênicos, 53 como não patogênicos e para 3 espécies não existem dados que comprovem sua patogenicidade, por isso foram classificados como patogenicidade indefinida. Dos 72 fungos patogênicos, 32 infectam plantas, 39 infectam animais e 1 infecta nematóides. Dos patógenos animais, 35 infectam vertebrados e 4 infectam somente insetos. Dos patógenos de vertebrados, todos infectam mamíferos, sendo que: 3 infectam mamíferos em geral, 30 infectam apenas

humanos, 1 infecta bovinos e humanos e 1 infecta tanto mamíferos quanto aves. Os dados da classificação estão contidos no Material Suplementar A.

3.8 Obtenção dos KOs significativamente enriquecidos nos genomas de fungos comparados aos demais eucariotos

3.8.1 Análises estatísticas por meio do CALANGO

O CALANGO (não publicado) é um programa que avalia elementos genômicos que estão significativamente super-representados em um grupo em relação a outro, sendo um dos grupos possuidores da característica ou fenômeno que desejo investigar e no outro grupo esta característica está ausente. O algoritmo do CALANGO executa com Bootstrap como método de reamostragem. Esse método realiza amostragens com reposição, em que as amostras são iguais aos dados originais. São feitos n sorteios com reposição da amostra inicial, em que n representa o número de observações disponíveis, e os dados reamostrados são amostras Bootstrap (FRANCO; REISEN, 2004). Esse procedimento foi executado 100 vezes no CALANGO para se construir uma distribuição e assim realizar inferências estatísticas.

As análises estatísticas do CALANGO foram feitas por meio do teste exato de Fisher, que é o tipo de teste utilizado para se investigar dados categóricos classificados de duas maneiras distintas em que se é analisada a significância da associação nos dois tipos de classificação (FISHER, 1922). Assim, ao se obter os dados de significância, é realizada uma correção FDR, do inglês *false discovery rate* que é um método utilizado para controle de erros do tipo 1 (rejeições incorretas da hipótese nula) (BENJAMINI; HOCHBERG, 1995).

Para executar o CALANGO são necessários 3 arquivos de INPUT: Teste, Background e Dicionário. O arquivo de Teste é composto pela lista dos IDs dos fungos e seus respectivos KOs (arquivo Fungi_ID_KO), o arquivo de Background é composto pela lista dos IDs dos eucariotos não-fungos e seus respectivos KOs (arquivo Eukanofungi_ID_KO), e o Dicionário é composto pela lista dos KEGG Orthology juntamente com a descrição de cada um (arquivo dicionario_eukaryotes). Inicialmente, é necessário editar os parâmetros do CALANGO, a fim de informar ao programa o caminho de cada um destes arquivos. Caso haja alguma discordância nos caminhos informados, o programa não é executado.

```
CALANGO <- list(type = "significance",
  output.dir = "results_FUNGxNFUNG/",
  test.path = "/home/alineis/mestrado/bin/CALANGO/bin/scripts/Fungi_ID_KO_teste_comarruma/",
  back.path = "/home/alineis/mestrado/bin/CALANGO/bin/scripts/Eukanofungi_ID_KO_background_comarruma/",
  ontology = "other",
  dict.path = "/home/alineis/mestrado/data/Ko_lists/dicionario_eukaryotes.txt",
  column = "KO",
  bootstrap = 100,
  criticalValue = 0.05,
  cores = 60)
```

Figura 11. Arquivo de parâmetros do CALANGO - Para que o programa seja executado, é necessário editar os parâmetros a fim de indicar os caminhos para cada arquivo de entrada.

Fonte: Dados da pesquisa

Para considerarmos um KEGG Orthology super-representado dentro do grupo dos fungos, este KO deveria atender aos seguintes critérios: possuir valor-q menor que 0.05, Bootstrap igual ou maior a 95% nas repetições e deveria possuir incidência zero no grupo de eucariotos não-fungos, ou seja, esse grupo de homólogos não deveria ocorrer em nenhum outro eucarioto não-fungo.

3.8.2 Obtenção das vias metabólicas KEGG e associação aos KOs

Assim como os dados dos genomas dos organismos e os KEGG Orthology, as vias metabólicas KEGG foram também obtidas por meio do KEGG Application Programming Interface e do protocolo REST. Desta forma, ao todo, fizemos o download de 463 vias metabólicas (KEGG Pathways).

wget http://rest.kegg.jp/list/pathway

```
path:map00010 Glycolysis / Gluconeogenesis
path:map00020 Citrate cycle (TCA cycle)
path:map00030 Pentose phosphate pathway
path:map00040 Pentose and glucuronate interconversions
path:map00051 Fructose and mannose metabolism
path:map00052 Galactose metabolism
path:map00053 Ascorbate and aldarate metabolism
path:map00061 Fatty acid biosynthesis
path:map00062 Fatty acid elongation
path:map00071 Fatty acid degradation
path:map00072 Synthesis and degradation of ketone bodies
path:map00073 Cutin, suberine and wax biosynthesis
path:map00100 Steroid biosynthesis
path:map00120 Primary bile acid biosynthesis
path:map00121 Secondary bile acid biosynthesis
path:map00130 Ubiquinone and other terpenoid-quinone biosynthesis
path:map00140 Steroid hormone biosynthesis
path:map00190 Oxidative phosphorylation
path:map00195 Photosynthesis
path:map00196 Photosynthesis - antenna proteins
path:map00220 Arginine biosynthesis
path:map00230 Purine metabolism
path:map00232 Caffeine metabolism
path:map00240 Pyrimidine metabolism
path:map00250 Alanine, aspartate and glutamate metabolism
path:map00253 Tetracycline biosynthesis
path:map00254 Aflatoxin biosynthesis
path:map00260 Glycine, serine and threonine metabolism
path:map00261 Monobactam biosynthesis
path:map00270 Cysteine and methionine metabolism
path:map00280 Valine, leucine and isoleucine degradation
path:map00281 Geraniol degradation
path:map00290 Valine, leucine and isoleucine biosynthesis
path:map00300 Lysine biosynthesis
path:map00310 Lysine degradation
path:map00311 Penicillin and cephalosporin biosynthesis
path:map00330 Arginine and proline metabolism
```

Figura 12. Arquivos das vias metabólicas KEGG - Na primeira coluna consta a lista das entradas KEGG Pathway e na segunda a descrição de cada uma delas.

Fonte: Dados da pesquisa

Em seguida, foi realizada uma edição do arquivo para retirarmos os caracteres que não são lidos pelo CALANGO e que o impediria de executar. Após isso, redirecionamos apenas a primeira coluna (correspondente aos IDs das vias metabólicas KEGG) para um novo arquivo:

```
for i in T*; do sed -i -e 's/path:./g' ${i} ;done > pathway_info_editado
```

```
for i in `less pathway_info_editado`; do awk '{print $1}'; done > ID_pathway
```

Seguidamente, esse novo arquivo, chamado ID_pathway, foi associado a todos os KOs relacionados a cada uma das vias KEGG. Para isso, foi realizado um download e link (via protocolo REST) da lista dos KEGG Orthology que ocorrem em cada via KEGG contida na nossa lista de vias. A saída do comando é direcionada para um novo arquivo executável que nos retorna uma lista de arquivos nomeados de acordo com o ID da via e seu conteúdo é a lista de todos os KOs associados a essa via:

```
for i in `less ID_pathway`; do echo wget http://rest.kegg.jp/link/ko/$i ; done > download_link_ko_pathway
```

map00010	K00011
map00020	K00035
map00030	K00094
map00040	K00704
map00051	K00844
map00052	K00845
map00053	K00849
map00061	K00850
map00062	K00883
map00071	K00917
map00072	K00963
map00073	K00965
map00100	K01084
map00120	K01182
map00121	K01187
map00130	K01189
map00140	K01190
map00190	K01193
map00195	K01203
map00196	K01220
map00220	K01229
map00230	K01631
map00232	K01635
map00240	K01684
map00250	K01784
map00253	K01785
map00254	K01819
map00260	K01835
map00261	K01854
map00270	K02079
map00280	K02080
map00281	K02744
map00290	K02745
map00300	K02746
map00310	K02747
map00311	K02773
map00330	K02774
ID_pathway	map00052 . KOs

Figura 13. Exemplo do conteúdo dos arquivos de vias KEGG gerados - Cada arquivo gerado na etapa de mapeamento dos KEGG Orthology possui uma lista dos KOs que estão descritos para tal via metabólica fúngica.

Fonte: Dados da pesquisa

3.8.3 Mapeamento dos KOs enriquecidos nas vias metabólicas KEGG

Após obter os arquivos das vias KEGG constando cada um de seus KOs mapeados, realizamos uma contagem dos mesmos. Em seguida, fizemos uma busca filtrando apenas as vias que continham KOs enriquecidos e redirecionando a saída do comando para um novo arquivo. É importante ressaltar que, para que o comando funcionasse, foi necessária uma edição dos arquivos para retirar os caracteres excedentes, como já feito anteriormente nos demais arquivos.

```
wc -l map* > num_KOs_por_pathway
```

```
for i in T*; do sed -i -e 's/path://g' ${i};done
```

```
for i in T*; do sed -i -e 's/ko://g' ${i};done
```

```
for i in `ls map*`; do grep -f KOs_FUN_NONFUN_enquecidos.txt $i >  
$i.comgrep; done
```

Posteriormente, os resultados foram ordenados de forma que, em uma coluna constasse a lista dos KOs daquela via, e na outra a contagem destes KOs em ordem decrescente, ou seja, do mais super-representado para o menos super-representado naquela via, facilitando assim a identificação da quantidade de KOs enriquecidos em cada via. Depois disso, reunimos os dados da contagem do total de KOs por pathway em um único arquivo a fim de calcular a razão de KOs enriquecidos por KOs totais em cada via metabólica KEGG.

```
wc -l *.comgrep | sort -nr > resultados_KOs_RICOS_FUNG_NONFUNG
```

```
less resultados_KOs_RICOS_FUNG_NONFUNG | awk '{print $2,$1}' | sort >  
resultados_KOs_RICOS_FUNG_NONFUNG_ordenado
```

```
wc -l * > quantidade_KOs_por_pathway
```

```
less quantidade_KOs_por_pathway | awk '{print $2,$1}' | sort >
quantidade_KOs_por_pathway_ordenado
```

map00010	98	402	total
map00020	58	66	map01100
map00030	83	36	map04111
map00040	79	35	map04113
map00051	108	26	map04011
map00052	77	24	map01110
map00053	50	15	map01130
map00061	39	14	map04138
map00062	24	10	map01120
map00071	57	8	map03013
map00072	8	8	map03008
map00073	14	6	map04141
map00100	34	6	map00513
map00120	18	5	map04139
map00121	15	5	map03040
map00130	58	4	map01212
map00140	48	4	map01200
map00190	213	4	map00909
map00195	63	4	map00630
map00196	42	4	map00520
map00220	59	4	map00051
map00230	208	4	map00040
map00232	16	3	map03450
map00240	99	3	map03440
map00250	70	3	map02020
map00253	17	3	map01040
map00254	13	3	map00901
map00260	99	3	map00680
map00261	28	3	map00514
map00270	115	3	map00403
map00280	69	3	map00400
map00281	15	3	map00380
map00290	19	3	map00360
map00300	45	3	map00254
map00310	87	2	map04213
map00311	15	2	map04120
map00330	107	2	map03460

Figura 14. Lista das vias metabólicas e a contagem de KOs enriquecidos em fungos - A primeira coluna de cada arquivo contém a lista das vias metabólicas fúngicas e a segunda contém o número de ocorrências de KOs enriquecidos na via fúngica em questão.

Fonte: Dados da pesquisa

Os dados obtidos foram transferidos para uma tabela no Excel e foram submetidos a dois tipos de estratégia de ordenação, a fim de explorar e entender melhor nossas análises. A primeira estratégia é a ordenação de acordo com a razão decrescente do número de KOs enriquecidos divididos pelo número de KOs totais presentes nas vias. A segunda estratégia é a ordenação de acordo com a quantidade decrescente do número total de KOs super-representados presentes nas vias.

3.9 Busca por chokepoints nas vias KEGG com KOs fúngicos enriquecidos

Na busca pelos chokepoints utilizamos como estratégia o mapeamento dos KOs enriquecidos nas vias utilizando um recurso do KEGG chamado *KEGG Mapper* - *Search&Color Pathway*. Esta ferramenta possibilita uma visualização dos objetos

KEGG (como por exemplo genes, proteínas, reações e compostos químicos) nas vias metabólicas nos quais estão presentes, sendo necessário fornecer a lista de identificadores KEGG do que se deseja mapear como entrada e são retornadas todas as vias às quais tais KOs pertencem. Ao se clicar nessas vias, o KEGG direciona para a página do mapa desta via e destaca a localização dos KOs inseridos como entrada, desta forma, os objetos encontrados são marcados em quaisquer cores (fornecidas juntamente com a lista de entrada).

KEGG Mapper – Search&Color Pathway

About KEGG Mapper

Reconstruct Pathway (and Brite, Module)

Search Pathway (and Brite, Module, Network, Disease)

Search&Color Pathway (and Brite, Module)

Color Pathway

Join Brite

Convert ID

Annotate Sequence

BlastKOALA

Map Taxonomy

KEGG

Target databases: Pathway, Brite hierarchy, Module

Search mode: ☒ Reference ☐ Organism-specific Enter: , ko, ec, rn, hsadd
Optional use of outside ID:

Enter objects one per line followed by bgcolor, fgcolor:

K02181	purple
K02307	purple
K02543	purple
K02544	purple
K02627	purple
K02646	purple
K03085	purple
K03356	purple
K06646	purple
K06647	purple

Examples:

Or upload file: Nenhum arquivo selecionado

If necessary, change default bgcolor:

☒ Include aliases

☐ Use uncolored diagrams

☐ Search pathways containing all the objects (AND search)

Figura 15. Exemplo de entrada no KEGG Mapper - A ferramenta KEGG Mapper possibilita a visualização de uma lista de componentes metabólicos de interesse do usuário dentro dos mapas KEGG.

Fonte: KEGG Mapper (modificado)

KEGG Mapper Search Result

Pathway (9)	Brite (9)	Module (0)
-------------	-----------	------------

Sort by the pathway list

Show matched objects

map04111 Cell cycle - yeast (36)

map04113 Meiosis - yeast (16)

map04011 MAPK signaling pathway

map04120 Ubiquitin mediated p

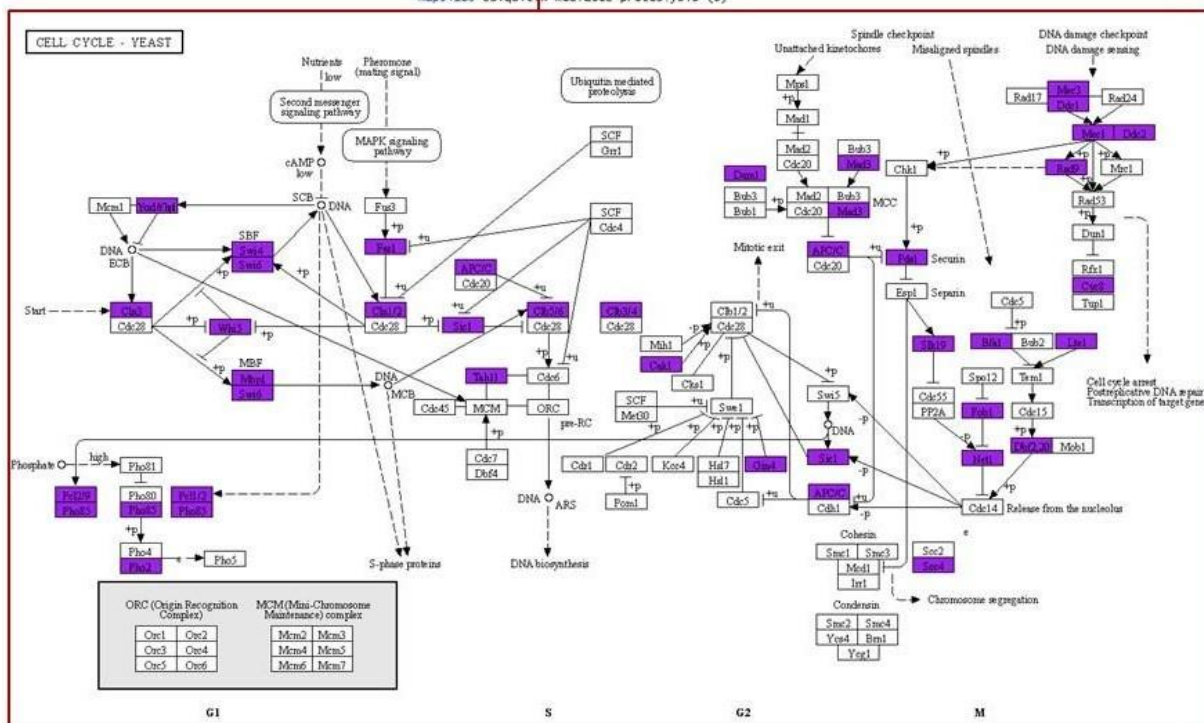


Figura 16. Exemplo de saída do KEGG Mapper - Após inserirmos o conjunto de KOs a serem mapeados juntamente com a cor desejada, o KEGG Mapper retorna a lista de vias metabólicas nas quais estes KOs são mapeados. Ao clicarmos em uma dessas vias, somos direcionados para o mapa da via em que estão destacados os KOs fornecidos que fazem parte da via.

Fonte: KEGG Mapper (modificado)

4 RESULTADOS

4.1 KOs mais significativamente super-representados em fungos em relação a não fungos

A saída do programa CALANGO é dada em forma de tabela que possui informações referentes aos KOs e à análise de enriquecimento dos mesmos, que são: 1) o identificador KO; 2) a descrição da função desse KO; 3) o valor-q considerando todos os genomas; 4) o número de bootstraps onde houve significância; 5) contagem de KOs encontrados no grupo de fungos; 6) contagem de KOs encontrados no grupo de eucariotos não-fungos.

Por conseguinte, transferimos esta tabela para o Excel, organizamos as colunas de acordo com os pontos de corte inicialmente estabelecidos e ordenamos a partir do valor-q, de modo a obtermos os KOs com maior diferença estatística ao compararmos fungos e não-fungos. Posteriormente, fizemos a contagem dos KOs filtrados obedecendo tais critérios, e encontramos 495 KOs super-representados no grupo dos fungos e ausentes no grupo dos demais eucariotos.

É importante ressaltar que uma discussão detalhada dos 495 KOs enriquecidos encontrados está fora do escopo do nosso trabalho e, por isso, comentaremos alguns exemplos dos KOs mais interessantes do ponto de vista da biologia de fungos. A seguir, a Tabela 1 mostra os 15 KOs mais super-representados no grupo dos fungos, de acordo com os critérios descritos anteriormente para significância. A tabela contendo todos os KOs enriquecidos encontra-se no material suplementar do presente trabalho.

Tabela 1. KOs enriquecidos ordenados por valor-q - KO ID = identificadores dos KEGG Orthology enriquecidos nos fungos; Annotation = descrições dos KOs; valor-q em ordem decrescente; valor de bootstrap; F = Fungos; NF = Não-Fungos.

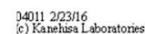
KO ID	ANNOTATION	qvalue	bootstrap (0.05)	F	NF
K19356	Lytic cellulose monooxygenase (C1-hydroxylating)	5,0582E-244	100	241	0
K19564	CTR, HNM1; choline transport protein	3,9285E-188	100	186	0
K09043	YAP; AP-1-like transcription factor	5,3171E-178	100	176	0
K08197	ARN; MFS transporter, SIT family, siderophore-iron:H ⁺ symporter	1,4441E-148	100	147	0
K19305	NPIL; deuterolysin	1,6515E-143	100	142	0
K11232	YPD1; osmolarity two-component system, phosphorelay intermediate protein YPD1	2,1969E-129	100	128	0
K17066	MOX; alcohol oxidase	2,2873E-126	100	125	0
K23787	DMA1_2; E3 ubiquitin-protein ligase DMA1/2	2,28E-124	100	123	0
K13682	BMT, WRY; beta-1,2-mannosyltransferase	2,3022E-123	100	122	0
K02143	ATPeFK, ATP19; F-type H ⁺ -transporting ATPase subunit k	2,2891E-121	100	120	0
K15564	CTK2; CTD kinase subunit beta	2,2576E-119	100	118	0
K17741	GRE2; NADPH-dependent methylglyoxal reductase	2,2503E-117	100	116	0
K02307	DAM1; DASH complex subunit DAM1	2,2052E-116	100	115	0
K17417	YMR31; small subunit ribosomal protein YMR-31	2,2695E-112	100	111	0
K15205	TFC6; transcription factor C subunit 6	2,3082E-111	100	110	0

Fonte: Dados da pesquisa

4.2 Vias metabólicas KEGG enriquecidas em KOs fúngicos

A segunda etapa da nossa pesquisa consistiu em mapear os KOs enriquecidos obtidos na etapa anterior nas vias KEGG às quais cada um destes KOs pertencem. Para exemplificar, trazemos nas figuras 17, 18 e 19 três das vias enriquecidas em vários KOs fúngicos: via map04011 - *MAPK signaling pathway (yeast)*, via da sinalização de MAPK em leveduras, map04111 - *Cell Cycle (Yeast)* e map04113 - *Meiosis (Yeast)*.

Nos exemplos a seguir, os KOs enriquecidos em fungos e ausentes nos demais eucariotos foram mapeados nas vias metabólicas em que são encontrados e estão destacados em verde (Figura 17), vermelho (Figura 18) e azul (Figura 19), indicando as etapas das quais participam.



Fonte: KEGG Mapper (modificado)

Fonte: KEGG Mapper (modificado)

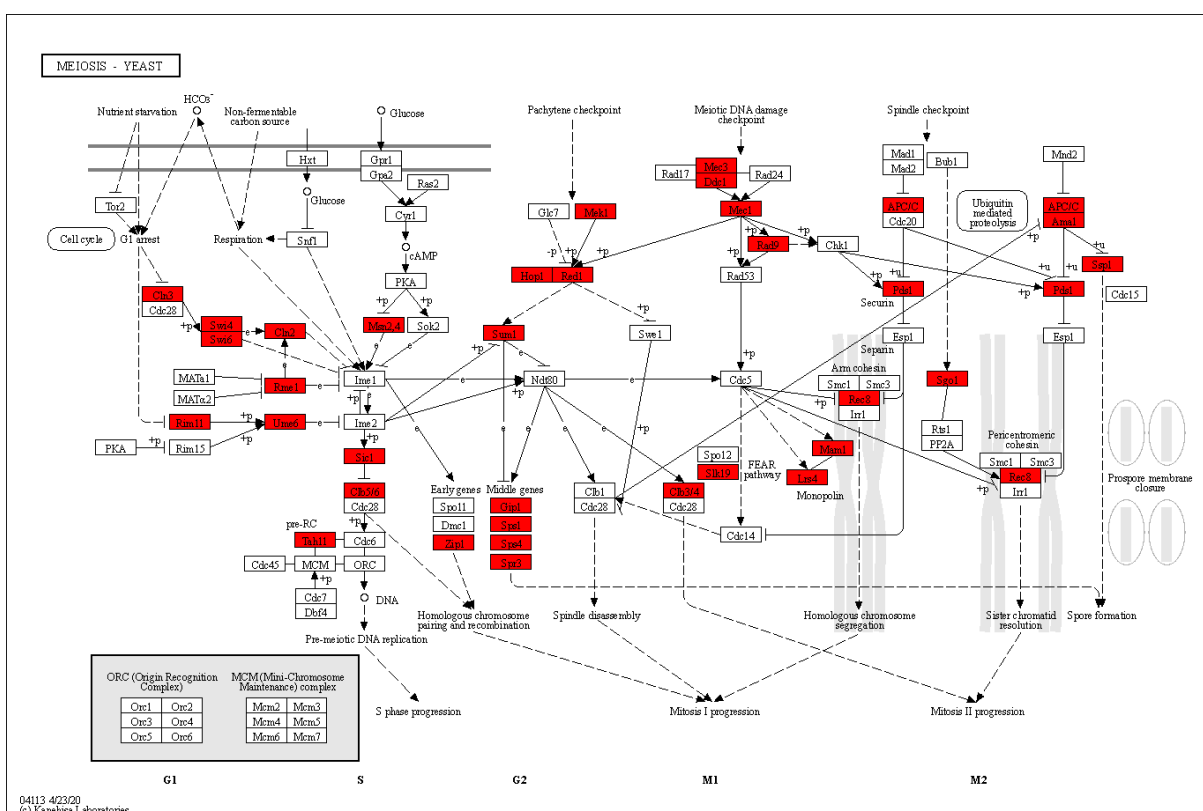


Figura 19. KOs enriquecidos em fungos mapeados na via metabólica map04113 - Meiosis (Yeast) - Em vermelho estão destacadas as reações enzimáticas codificadas por pelo menos um KO

que é enriquecido em fungos e ausente nos demais eucariotos na via da sinalização de meiose em leveduras.

Fonte: KEGG Mapper (modificado)

O mapeamento dos KOs enriquecidos nas vias KEGG resultou em uma tabela que organizamos conforme duas estratégias, como já mencionado. A primeira foi a ordenação de acordo com a razão entre o total de KOs enriquecidos nos fungos e os KOs totais mapeados nas vias metabólicas, sendo mais próxima de 1 a razão, maior o número de KOs enriquecidos em fungos que estão presentes na via quando comparado ao total de KOs da mesma via.

Tabela 2. Vias metabólicas KEGG enriquecidas em KOs fúngicos – Ordenadas pela razão de KOs enriquecidos nos fungos - PATHWAY = identificadores das vias metabólicas KEGG que contém KOs enriquecidos nos fungos; Annotation = descrições das vias metabólicas; Enriched KOs = número de KOs enriquecidos em fungos mapeados na via; Total KOs = total de KOs mapeados na via; RATIO = razão de KOs enriquecidos dividido pelo número total de KOs mapeados na via KEGG.

PATHWAY	ANNOTATION	ENRICHED KOs	TOTAL KOs	RATIO
map04113	Meiosis - yeast	35	103	0,340
map04111	Cell cycle - yeast	36	126	0,286
map04011	MAPK signaling pathway - yeast	26	91	0,286
map00254	Aflatoxin biosynthesis	3	13	0,231
map00403	Indole diterpene alkaloid biosynthesis	3	13	0,231
map04138	Autophagy - yeast	14	77	0,182
map03450	Non-homologous end-joining	3	19	0,158
map04139	Mitophagy - yeast	5	34	0,147
map00791	Atrazine degradation	2	16	0,125
map00513	Various types of N-glycan biosynthesis	6	52	0,115
map00901	Indole alkaloid biosynthesis	3	30	0,100
map03008	Ribosome biogenesis in eukaryotes	8	82	0,098
map00514	Other types of O-glycan biosynthesis	3	31	0,097
map01040	Biosynthesis of unsaturated fatty acids	3	32	0,094
map00785	Lipoic acid metabolism	1	11	0,091
...	...	...	...	...
Total		402	40713	

Fonte: Dados da pesquisa

Tabela 3. Vias metabólicas KEGG enriquecidas em KOs fúngicos – Ordenadas pela quantidade decrescente de KOs enriquecidos nos fungos - PATHWAY = identificadores das vias metabólicas KEGG que contém KOs enriquecidos nos fungos; Annotation = descrições das vias metabólicas; Enriched KOs = número de KOs enriquecidos em fungos mapeados na via; Total KOs = total de KOs mapeados na via; RATIO = razão de KOs enriquecidos dividido pelo número total de KOs mapeados na via KEGG.

PATHWAY	ANNOTATION	ENRICHED KOs	TOTAL KOs	RATIO
map01100	Metabolic pathways	66	4187	0,016
map04111	Cell cycle - yeast	36	126	0,286
map04113	Meiosis - yeast	35	103	0,340
map04011	MAPK signaling pathway - yeast	26	91	0,286
map01110	Biosynthesis of secondary metabolites	24	1885	0,013
map01130	Biosynthesis of antibiotics	15	1015	0,015
map04138	Autophagy - yeast	14	77	0,182
map01120	Microbial metabolism in diverse environments	10	1180	0,008
map03008	Ribosome biogenesis in eukaryotes	8	82	0,098
map03013	RNA transport	8	138	0,058
map00513	Various types of N-glycan biosynthesis	6	52	0,115
map04141	Protein processing in endoplasmic reticulum	6	141	0,043
map04139	Mitophagy - yeast	5	34	0,147
map03040	Spliceosome	5	122	0,041
map00909	Sesquiterpenoid and triterpenoid biosynthesis	4	73	0,055
...	...	...	...	...
Total		402	40713	

Fonte: Dados da pesquisa

Ao todo, foram mapeados 402 KOs super-representados em fungos em pelo menos uma via KEGG, o que representa 81,21% dos KOs enriquecidos obtidos na primeira etapa da nossa pesquisa. Nosso mapeamento retornou 91 vias KEGG (~20% do total de vias KEGG) contendo pelo menos 1 KO enriquecido. Novamente, a discussão detalhada de cada uma destas vias foge do escopo do presente trabalho. Assim, disponibilizamos no corpo do presente documento as 15 vias metabólicas mais enriquecidas ordenadas por razão de KOs (Tabela 2) e por total absoluto (Tabela 3) de KOs enriquecidos nos grupos de genes homólogos fúngicos.

4.3 Candidatos a chokepoints nas vias KEGG contendo KOs fúngicos enriquecidos

Dentre as vias contendo KOs super representados em fungos, realizamos a busca, via curadoria manual da vias KEGG, por possíveis chokepoints nas vias responsáveis pela biossíntese de produtos celulares e que, ao terem alguma etapa

inibida, poderiam acarretar em algum dano ao organismo (DOERING, 1999).

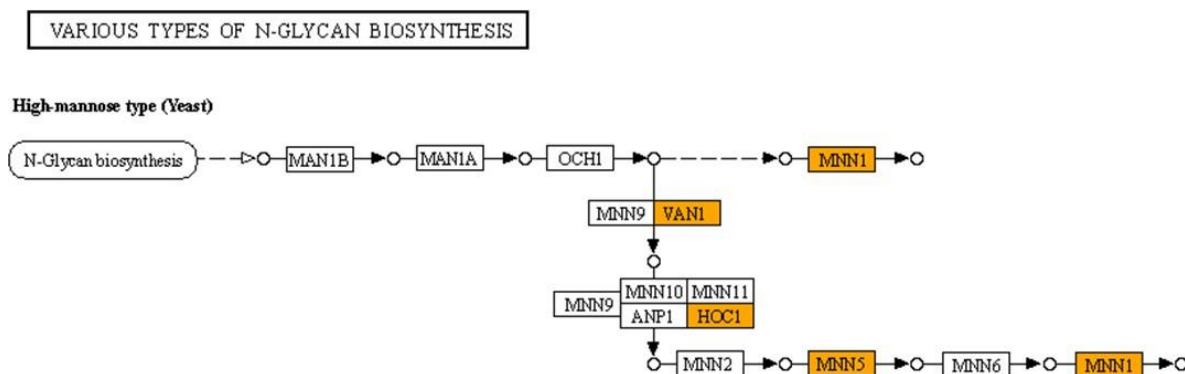


Figura 20. Via da Biossíntese de vários tipos de N-Glicanos com os KOs fúngicos enriquecidos em destaque - Em laranja estão destacadas as reações enzimáticas codificadas por pelo menos um KO que é enriquecido em fungos e ausente nos demais eucariotos na via da biossíntese de N-glicanos. Fonte: KEGG Mapper (modificado)

Desta forma, encontramos dois candidatos interessantes a *chokepoints*, que fazem parte da via de síntese de N-Glicanos, que se encaixam na definição de um *chokepoint*. São eles as enzimas MNN1 (alfa-1,3-manosiltransferase) e MNN5 (alfa-1,2-manosiltransferase) (figura 20).

5 DISCUSSÃO

5.1 KOs mais significativamente super-representados em fungos em relação a não fungos

Do total de 13962 KOs encontrados em pelo menos um genoma eucariótico, descobrimos que 495 são significativamente enriquecidos no grupo dos fungos (F). Entre os 50 KOs mais abundantes, encontramos marcas claras de suas funções ecológicas como decompositores e atividades metabólicas de interesse comercial. Um exemplo é o KO K19356 - *lytic cellulose monooxygenase (C1-hydroxylating)* [EC:1.14.99.54], uma enzima da via de degradação de celulose muito estudada devido às suas propriedades oxidativas que permitem a otimização do processo de deterioração da celulose, o que a torna um alvo interessante para aplicação no campo biotecnológico, como por exemplo, aumentando a produtividade no processo de produção de biocombustíveis (BEESON et al., 2012; GARAJOVA et al., 2016; VILLARES et al., 2017).

Encontramos também um KO que está relacionado a atividades metabólicas aproveitadas comercialmente (K19305 - NP11; *Deuterolysin* [EC:3.4.24.39]). Podemos exemplificar com o estudo de Maeda et al. (2016) que afirma que a Deuterolisina é uma enzima encontrada em fungos como *Aspergillus oryzae* e *Aspergillus sojae*, estes organismos são utilizados na produção de alimentos fermentados no Japão e são classificados como seguros pela *Drug Food Association*. Em contrapartida, Li et al. (2012), por meio de uma análise de genômica comparativa, propõe que houve um evento de seleção positiva que levou a uma expansão dos genes da família da Deuterolisina nas espécies de fungos patogênicos *Coccidioides posadasii* e *Coccidioides immitis*. Tal estudo sugere, portanto, a associação entre essa expansão da família de genes da Deuterolisina eo desenvolvimento de patogenicidade nessas espécies ao longo de sua evolução.

Ademais, obtivemos um KO já descrito na literatura como estando envolvido no processo de resistência a múltiplas drogas - K08197 - *MFS transporter, SIT family, siderophore-iron: H⁺ symporter*. A sigla MFS em português significa “Superfamília dos Facilitadores Maioritários”, que são proteínas de membrana que desempenham um importante papel na aquisição da resistência à antifúngicos em

leveduras (DOS SANTOS et al., 2014; GULSHAN; MOYE-ROWLEY, 2007; SÁ-CORREIA et al., 2009).

Interessantemente, nossa estratégia também selecionou enzimas já utilizadas como alvos de antifúngicos, tais como o K23787 (DMA1_2; E3 *ubiquitin-protein ligase* DMA1/2 [EC:2.3.2.23]) (ZHU et al., 2016; VAN DE VEERDONK; NETEA, 2016; LIANG et al., 2020; QIN et al., 2019) e o KO K19564 (CTR, HNM1; *choline transport protein*) (ROGGERO et al., 2004). Esse fato sugere que nossa estratégia para a seleção de genes enriquecidos no genoma de fungos foi capaz de encontrar alvos de medicamentos anti-fúngicos já descritos por outros pesquisadores.

Finalmente, encontramos vários candidatos interessantes para pesquisas futuras que desempenham um papel importante no estilo de vida dos fungos, como reguladores de transcrição e componentes de processos de parede celular. Um regulador de transcrição interessante que encontramos é o K09043 - YAP; *AP-1-like transcription factor*, que possui função já conhecida em fungos como regulador de resposta ao estresse oxidativo, incluindo as que são causadas por metais e drogas (FERNANDES; RODRIGUES-POUSADA; STRUHL, 1997; LEE et al., 1999; RODRIGUES-POUSADA et al., 2019). Alguns estudos propõem que o fator de transcrição AP-1-Like está relacionado à virulência em alguns fungos patogênicos, como *Candida Albicans*, *Cryptococcus neoformans* e *Talaromyces marneffe* (*Penicillium marneffe*) (PATTERSON et al., 2013; SIMAAN et al., 2020; SO et al., 2019).

Exemplificando um componente importante de processos de parede celular dentre os KOs enriquecidos encontrados, destacamos o K13682 (BMT, WRY; *beta-1,2-mannosyltransferase* [EC:2.4.1.-]). A composição da parede celular dos fungos confere uma das características específicas desse Reino, já que a composição básica da mesma é a quitina, composto encontrado na parede celular dos fungos. Além disso, a enzima *beta-1,2-mannosyltransferase* foi abordada em estudos que destacam sua associação a fungos patogênicos como *Candida Albicans* e *Aspergillus fumigatus* (COURJOL et al., 2015; HALL et al., 2013; MILLE et al., 2008; ONOUE et al., 2018).

5.2 Vias metabólicas KEGG enriquecidas em KOs fúngicos em relação aos eucariontes não-fungos

Nas Tabelas 2 e 3 utilizamos dois critérios distintos de ordenação (a quantidade decrescente relativa e a quantidade decrescente absoluta de KOs encontrados por via, respectivamente). O que podemos observar analisando ambas tabelas de vias enriquecidas é que, independentemente da estratégia que usamos para ordenação, encontramos diversas vias exclusivas nos fungos entre os 15 primeiros lugares do nosso ranking, embora a estratégia de ordenação por razão parece recuperar mais via fungo-específicas.

Entre as 15 vias que contêm as maiores frações de KOs de fungos enriquecidos, 7 são vias KEGG específicas para fungos: map04113 - Meiosis (yeast); map04111 - Cell cycle (yeast); map04011 - MAPK signaling pathway (yeast); map04138 - Autophagy (yeast); map04139 - *Mitophagy* (yeast), que são vias ligadas à processos metabólicos fundamentais à sobrevivência celular; e as vias map00254 - *Aflatoxin biosynthesis*; map00403 - *Indole diterpene alkaloid biosynthesis*, que são vias de produção de toxinas exclusivas de fungos (CLIFT; BIZZARI; MARSTON, 2009; DIAMOND et al., 2009; KASSIR et al., 2003; SOURIRAJAN; LICHTEN, 2008; CHEN; ZHOU, 2009; COSTANZO et al., 2004; DE BRUIN et al., 2004; DEVAULT et al., 2002; DIRICK; BÖHM; NASMYTH, 1995; FU; PASTUSHOK; XIAO, 2008; SIMANIS, 2003; HE; KLIONSKY, 2009; NAKATOGAWA et al., 2009; POLLACK; HARRIS; MARTEN, 2009; REGGIORI; KLIONSKY, 2013; EHRLICH, 2009; KORMAN et al., 2010; SAKUNO et al., 2005; SHIMA et al., 2009; YABE; NAKAJIMA, 2004; YU, 2012; MOTOYAMA et al., 2012; NICHOLSON et al., 2009; SAIKIA et al., 2006, 2007, 2008, 2012).

Vale destacar a via de sinalização de MAPK em fungos, uma vez que essa via está diretamente ligada a vários processos importantes para a viabilidade fúngica, tais como a regulação da integridade da parede celular e fatores de transcrição reguladores da expressão gênica. Essa via também está ligada à patogênese dos fungos em plantas por espécies como *Fusarium graminearum*, *Cochliobolus sativus* e *Mycosphaerella graminicola*. Conjuntamente, conclui-se que essa pode ser uma via pertinente de se estudar como alvo de fungicidas para fitopatógenos (JIANG et al., 2018; LIN et al., 2010; MARTÍNEZ-SOTO; RUIZ-HERRERA, 2017).

Um ponto relevante quanto à via exclusiva de autofagia em leveduras (map04138) é que a mesma foi associada por vários autores aos processos relacionados à patogenicidade, tanto em fungos filamentosos quanto em leveduras, o que é interessante a se considerar quando se pensa em busca de novos alvos antifúngicos (KHAN et al., 2012; LIU et al., 2012, 2016; PALMER; ASKEW; WILLIAMSON, 2008; ZHU et al., 2019). Outro fato interessante dos nossos resultados é que ambas as vias map04113 - *Meiosis (yeast)* e map04111 - *Cell cycle (yeast)* estão relacionadas à via da proteólise mediada pela proteína ubiquitina (HENCHOZ et al., 1997; PESIN; ORR-WEAVER, 2008; SKOWYRA et al., 1999), uma vez que a enzima responsável pela ligação da ubiquitina a outras proteínas celulares foi encontrada como super-representada na primeira etapa dos nossos resultados (K23787 - DMA1_2; E3 *ubiquitin-protein ligase*) que como mencionamos é uma das proteínas já utilizadas como alvo de drogas antifúngicas (ZHU et al., 2016; VAN DE VEERDONK; NETEA, 2016; LIANG et al., 2020; QIN et al., 2019).

Dentre nossos achados, também obtivemos 4 vias que, embora não sejam fungo-específicas, são observadas quase exclusivamente nesse táxon (vias observadas em até cinco outros organismos, apenas). Duas destas vias estão diretamente ligadas a composição de parede celular fúngica (map00513 - *Various types of N-glycan biosynthesis* e map00514 - *Other types of O-glycan biosynthesis*), enquanto uma das duas vias restantes são responsáveis pela produção de toxinas (map00901 - *Indole alkaloid biosynthesis*), enquanto a segunda é a via responsável pela degradação da Atrazina (map00791 - *Atrazine degradation*), que é um composto químico utilizado como herbicida para controle de ervas daninhas em plantações de milho, cana-de açúcar e sorgo (CHAN-CUPUL; HEREDIA-ABARCA; RODRÍGUEZ-VÁZQUEZ, 2016; ELGUETA et al., 2016; FUKUDA et al., 2009; GEERLINGS et al., 2000; MARINHO et al., 2017; RIJHWANI; SHANKS, 1998).

Também encontramos diversas vias filogeneticamente diversas de processos celulares basais que contêm consideráveis frações de KOs enriquecidos em genomas de fungos, tais como a via de reparo de DNA (map03450 - *Non-homologous end-joining*), síntese de proteínas (map03008 - *Ribosome biogenesis in eukaryotes*) e metabolismo de ácidos graxos (map01040 - *Biosynthesis of unsaturated fatty acids*). Como essas vias representam processos primordiais para viabilidade celular, estes KOs, podem compreender objetivos interessantes para uma intervenção mais sistêmica na biologia de fungos como, por

exemplo, na busca por *chokepoints* (SINGH; MALIK; SHARMA, 2007; TAYLOR et al., 2013).

5.3 Candidatos a *chokepoints* nas vias KEGG contendo KOs fúngicos enriquecidos

Após extensa curadoria manual das vias KEGG contendo KOs enriquecidos em fungos, encontramos um candidato a *chokepoint* que é também um componente da via de síntese de parede celular em fungos. A enzima MNN1 (K05538) representa o gene codificador de proteínas YER001W. A enzima MNN1 é uma das cinco proteínas de *Saccharomyces cerevisiae* da família multigênica MNN, e trata-se de uma glicoproteína de membrana integral do complexo de Golgi necessária para a adição de ligações alfa-1,3-manose a oligossacarídeos ligados a N (nitrogênio) e O (oxigênio). Em leveduras, na via da biossíntese de N-glicanos, o objetivo desta enzima nesta via é adicionar manoses ao glicano G10841, convertendo-o em G11040. Da mesma forma acontece com o glicano G13049 que é convertido em G13050, conforme mostra a figura 20.

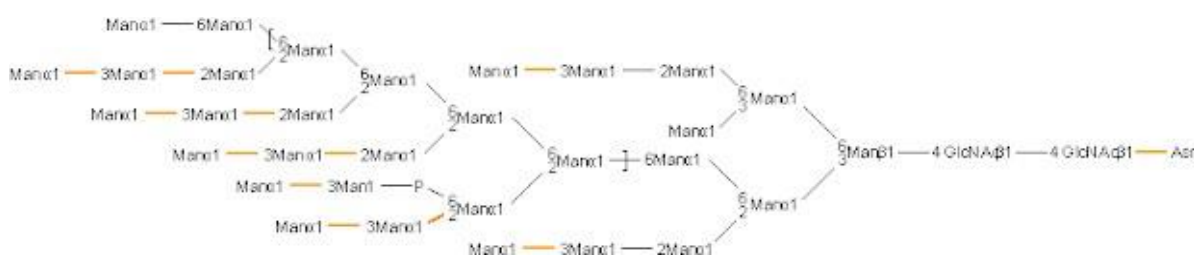


Figura 21. N-glicano com alto teor de manose com destaque para as os KOs fúngicos enriquecidos - Representação do KEGG mapper das ligações (destacadas em amarelo) da molécula complexa das quais as enzimas MNN1 e MNN5 participam na via metabólica de biossíntese de N-glicanos.

Fonte: KEGG Mapper (modificado)

A enzima MNN5 (K05536) em *Saccharomyces cerevisiae* representa o gene codificador de proteínas YJL186W. Este gene possui função similar ao do MNN1 e da mesma forma, através da adição de manose converte o glicano G13047 em G13048. Ambas enzimas são necessárias para a conformidade estrutural de N-glicano com alto teor de manose (*High-mannose N-glican*) (figura 21), como a membrana do complexo de Golgi e parede celular de leveduras (LUSSIER; BUSSEY, 1999; KLUTTS, 2006; ZHANG et al., 2016; ONOUE et al., 2018).

As glicosiltransferases são um grupo de enzimas que nos fungos formam a quitina, principal composto da parede celular desses organismos. Dentro deste grupo temos as manosiltransferases, que são um grupo de enzimas interessantes, já são conhecidamente atribuídas a funções de adesão entre as células, habilidades imunomoduladoras e variabilidade antigênica (YIP et al., 1994; LIZICAROVA et al, 2007).

Entre 60 e 90% do peso seco da parede celular de *S. cerevisiae* é constituída de glicanos e manoses, o que significa que ao se inibir vias de síntese de componentes de parede celular, é possível inviabilizar a célula ou a deixar mais suscetível à ação das células de defesa do hospedeiro (ZHANG et al., 2016).

6 CONCLUSÕES

Nossa pesquisa realizou a análise de genômica comparativa de fungos versus eucariotos não-fungos em 2 etapas: em nível de genes homólogos e de vias KEGG. No nível de genes (KOs), obtivemos aproximadamente 400 KOs enriquecidos em fungos. Dentre os 50 mais super-representados, encontramos genes intimamente ligados à função dos fungos como decompositores, bem como às suas atividades de interesse comercial e biotecnológica, além de enzimas já utilizadas como alvos de antifúngicos e de vários candidatos interessantes para pesquisas futuras que desempenham um papel importante no estilo de vida dos fungos, como reguladores de transcrição e componentes de processos de parede celular.

No nível de vias, entre as vias que contém o maior número de KOs enriquecidos, encontramos tanto vias que são específicas dos fungos quanto módulos conservados em vias mais abrangentes filogeneticamente. Consideramos que a detecção das vias fungo-específicas como um importante controle positivo da nossa estratégia, uma vez que um dos nossos objetivos iniciais era detectar vias enriquecidas nos fungos. Uma vez que nossa estratégia foi capaz de encontrar vias que são exclusivas em tais organismos, consideramos esse sinal claro de que a nossa estratégia foi satisfatória.

Algumas das vias enriquecidas específicas nos fungos compreendem processos celulares fundamentais que promovem a sobrevivência celular dos fungos (inclusive sob estresse ambiental). Além da detecção destas vias enriquecidas específicas, nossa estratégia também detectou vias ligadas a processos essenciais para a vida celular tais como reparo de DNA, síntese de proteínas e metabolismo de ácidos graxos, que podem compreender alvos para estudos de novas proteínas drogáveis ou de *chokepoints*.

Adicionalmente, nosso trabalho também foi capaz de selecionar potenciais novos alvos para o desenvolvimento de drogas anti-fúngicas através da análise de *chokepoints*. Assim, acreditamos que este trabalho possui uma metodologia capaz de contribuir na compreensão do *modus operandi* de organismos complexos como os fungos.

7 PERSPECTIVAS

Apresentamos um método de análise que pode constituir uma interessante ferramenta de genômica comparativa em nível de genes e vias metabólicas na descoberta de padrões moleculares específicos de um grupo de organismos de interesse científico em relação a outro (por exemplo, no estudo comparativo entre espécies de plantas, protozoários, etc.). Também é possível utilizar a mesma estratégia para caracterizar grupos de organismos fenotipicamente similares (por exemplo, patógenos de uma mesma espécie *versus* não patógenos).

Nossos resultados apresentaram potenciais novos alvos para busca por *chokepoints*, como também enzimas exclusivas em vias metabólicas fúngicas que conceitualmente se constituem como *chokepoints*. Novos trabalhos são necessários tanto para uma investigação mais extensa dos dados obtidos, quanto para a validação experimental destes candidatos.

REFERÊNCIAS

- AHRENDT, S. R. et al. Leveraging single-cell genomics to expand the fungal tree of life. **Nature Microbiology**, v. 3, n. 12, p. 1417–1428, 2018.
- ALTENHOFF, A. M.; GLOVER, N. M.; DESSIMOZ, C. Inferring Orthology and Paralogy BT - Evolutionary Genomics: Statistical and Computational Methods. In: ANISIMOVA, M. (Ed.). New York, NY: **Springer** New York, 2019. p. 149–175.
- AOKI-KINOSHITA, K.F.; KANEHISA, M. Gene annotation and pathway mapping in KEGG. **Methods Mol Biol.** 2007;396:71-91. doi: 10.1007/978-1-59745-515-2_6. PMID: 18025687.
- ARAÚJO, R; SAMPAIO-MAIA, B. Fungal Genomes and Genotyping. **Adv Appl Microbiol.** 2018. 102:37-81. doi: 10.1016/bs.aambs.2017.10.003. Epub 2017 Dec 7. PMID: 29680126.
- CASADEVALL, Arturo. Fungal Diseases in the 21st Century: The Near and Far Horizons. **Pathog Immun.** 2018;3(2):183-196. doi: 10.20411/pai.v3i2.249. Epub 2018 Sep 25. PMID: 30465032; PMCID: PMC6241320.
- BALDAUF, S. L.; PALMER, J. D. Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. **Proceedings of the National Academy of Sciences of the United States of America**, v. 90, n. 24, p. 11558–11562, 1993.
- BARKE, J.; SEIPKE, R. F.; GRÜSCHOW, S.; HEAVENS, D.; DROU, N.; BIBB, M. J.; GOSS, R. J.; YU, D. W.; HUTCHINGS, M. I. (2010). A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant *Acromyrmex octospinosus*. **BMC biology**, 8, 109. <https://doi.org/10.1186/1741-7007-8-109>
- BAYLY, A. M.; MACREADIE, I. G. Folic acid antagonism of sulfa drug treatments. **Trends Parasitol** 18, 49–50 (2002).
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 57, n. 1, p. 289–300, jan. 1995.
- BEESON, W. T. et al. Oxidative Cleavage of Cellulose by Fungal Copper-Dependent Polysaccharide Monooxygenases. **Journal of the American Chemical Society**, v. 134, n. 2, p. 890–892, 18 jan. 2012.
- BRUNS, T. A kingdom revised. **Nature**, v. 443, n. 7113, p. 758–761, 18 out. 2006.
- BUCKLEY, M. The Fungal Kingdom: diverse and essential roles in earth's ecosystem: This report is based on a colloquium, sponsored by the American Academy of Microbiology, convened November 2–4, 2007 in Tucson, Arizona. Washington (DC): **American Society for Microbiology**; 2008. PMID: 32687285.
- CHAKRAVORTY, S.; HELB, D.; BURDAY, M.; CONNELL, N; ALLAND, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. **J Microbiol Methods.** 2007 May;69(2):330-9. doi: 10.1016/j.mimet.2007.02.005. Epub 2007 Feb 22. PMID: 17391789; PMCID: PMC2562909.

CHAN-CUPUL, W.; HEREDIA-ABARCA, G.; RODRÍGUEZ-VÁZQUEZ, R. Atrazine degradation by fungal co-culture enzyme extracts under different soil conditions.

Journal of Environmental Science and Health - Part B Pesticides, Food Contaminants, and Agricultural Wastes, v. 51, n. 5, p. 298–308, 2016.

CHEN, S. H.; ZHOU, H. Reconstitution of Rad53 activation by Mec1 through adaptor protein Mrc1. **Journal of Biological Chemistry**, v. 284, n. 28, p. 18593–18604, 2009.

CLIFT, D.; BIZZARI, F.; MARSTON, A. L. Shugoshin prevents cohesin cleavage by PP2A^{Cdc55}-dependent inhibition of separase. **Genes and Development**, v. 23, n. 6, p. 766–790, 2009.

COSTANZO, M. et al. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. **Cell**, v. 117, n. 7, p. 899–913, 2004.

COURJOL, F. et al. β -1,2-Mannosyltransferases 1 and 3 Participate in Yeast and Hyphae O- and N-Linked Mannosylation and Alter *Candida albicans* Fitness During Infection. **Open Forum Infectious Diseases**, v. 2, n. 3, p. 2633851, 1 set. 2015.

DE BRUIN, R. A. M. et al. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. **Cell**, v. 117, n. 7, p. 887–898, 2004.

DEVAULT, A. et al. Identification of Tah11/Sid2 as the ortholog of the replication licensing factor Cdt1 in *Saccharomyces cerevisiae*. **Current Biology**, v. 12, n. 8, p. 689–694, 2002.

DIAMOND, A. E. et al. The Anaphase Promoting Complex Targeting Subunit Ama1 Links Meiotic Exit to Cytokinesis during Sporulation in *Saccharomyces cerevisiae*. **Molecular Biology of the Cell**, v. 20, n. 1, p. 134–145, jan. 2009.

DIRICK, L.; BÖHM, T.; NASMYTH, K. Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of *Saccharomyces cerevisiae*. **EMBO Journal**, v. 14, n. 19, p. 4803–4813, 1995.

DOERING, T. L. A Unique α -1,3 Mannosyltransferase of the Pathogenic Fungus *Cryptococcus neoformans*. **Journal of Bacteriology**, v. 181, n. 17, p. 5482–5488, 1 set. 1999.

DOS SANTOS, S. C. et al. MFS transporters required for multidrug/multixenobiotic (MD/MX) resistance in the model yeast: Understanding their physiological function through post-genomic approaches. **Frontiers in Physiology**, v. 5 MAY, n. May, p. 1–15, 2014.

DUNHAM, M.J.; BADRANE, H.; FEREA, T.; ADAMS, J.; BROWN, P.O.; ROSENZWEIG, F.; BOTSTEIN, D. (2002). Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. **Proceedings of the National Academy of Sciences** Dec 2002, 99 (25) 16144-16149. DOI: 10.1073/pnas.242624799

EHRLICH, K. C. Predicted roles of the uncharacterized clustered genes in aflatoxin biosynthesis. **Toxins**, v. 1, n. 1, p. 37–58, 2009.

ELGUETA, S. et al. Immobilization of the white-rot fungus *Anthracoophyllum discolor* to degrade the herbicide atrazine. **AMB Express**, v. 6, n. 1, 2016.

FERNANDES, L.; RODRIGUES-POUSADA, C.; STRUHL, K. Yap, a novel family of eight bZIP proteins in *Saccharomyces cerevisiae* with distinct biological functions. **Molecular and Cellular Biology**, v. 17, n. 12, p. 6982–6993, 1997.

FIELDING, Roy Thomas. **Architectural Styles and the Design of Network-based Software Architectures**. Doctoral dissertation, University of California, Irvine, 2000.

FISHER, R. A. “On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P.” **Journal of the Royal Statistical Society**, vol. 85, no. 1, 1922, pp. 87–94. JSTOR, www.jstor.org/stable/2340521.

FITCH, W. M. Distinguishing Homologous from Analogous Proteins. **Systematic Biology**, v. 19, n. 2, p. 99–113, 1 jun. 1970.

FITCH, W. M. Homology: a personal view on some of the problems. **Brenner's Encyclopedia of Genetics: Second Edition**, v. 16, p. 227–231, 2000.

FRANCO, G. C.; REISEN, V. A. Bootstrap techniques in semiparametric estimation methods for ARFIMA models: A comparison study. **Computational Statistics**, v. 19, n. 2, p. 243–259, 1 maio 2004.

FU, Y.; PASTUSHOK, L.; XIAO, W. DNA damage-induced gene expression in *Saccharomyces cerevisiae*. **FEMS Microbiology Reviews**, v. 32, n. 6, p. 908–926, 2008.

FUENTEFRÍA, A. M. et al. Antifungals discovery: an insight into new strategies to combat antifungal resistance. **Letters in Applied Microbiology**, v. 66, n. 1, p. 2–13, jan. 2018.

FUKUDA, E. K. et al. Fungal cell wall polysaccharides: Purification and characterization. **Semina: Ciências Agrárias**, v. 30, n. 1, p. 117–134, 2009.

GALAGAN, J. E. et al. Genomics of the fungal kingdom: Insights into eukaryotic biology. **Genome Research**, v. 15, n. 12, p. 1620–1631, 2005.

GARAJOVA, S. et al. Single-domain flavoenzymes trigger lytic polysaccharide monooxygenases for oxidative degradation of cellulose. **Scientific Reports**, v. 6, n. March, p. 1–9, 2016.

GEERLINGS, A. et al. Molecular cloning and analysis of strictosidine β -D-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. **Journal of Biological Chemistry**, v. 275, n. 5, p. 3051–3056, 2000.

GOFFEAU, A.; BARRELL, B.G.; BUSSEY, H.; DAVIS, R.W.; DUJON, B.; FELDMANN H.; GALIBERT F.; HOHEISEL, J.D.; JACQ C.; JOHNSTON, M.; LOUIS

E.J.; MEWES, H.W.; MURAKAMI Y.; PHILIPPSEN, P; TETTELIN, H.; OLIVER, S.G. Life with 6000 genes. **Science**. 1996 Oct 25;274(5287):546, 563-7. doi: 10.1126/science.274.5287.546. PMID: 8849441.

GRIGORIEV, I. V.; NIKITIN, R.; HARIDAS, S.; KUO, A.; OHM, R.; OTILLAR, R.; RILEY, R.; SALAMOV, A.; ZHAO, X.; KORZENIEWSKI, F.; SMIRNOVA, T.; NORDBERG, H.; DUBCHAK, I.; SHABALOV, I. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. **Nucleic acids research**, 42(Database issue), D699–D704. <https://doi.org/10.1093/nar/gkt1183>

GRÜTZMANN, K.; SZAFRANSKI, K.; POHL, M.; VOIGT, K.; PETZOLD, A.; SCHUSTER, S. (2014). Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. **DNA Res.** 21, 27–39. 10.1093/dnares/dst038

GULSHAN, K.; MOYE-ROWLEY, W. S. Multidrug resistance in fungi. **Eukaryotic Cell**, v. 6, n. 11, p. 1933–1942, 2007.

HALL, R. A. et al. The Mnn2 Mannosyltransferase Family Modulates Mannoprotein Fibril Length, Immune Recognition and Virulence of *Candida albicans*. **PLoS Pathogens**, v. 9, n. 4, p. 13–17, 2013.

HAWKSWORTH, D.L. Global species numbers of fungi: Are tropical studies and molecular approaches contributing to a more robust estimate? **Biodivers. Conserv.** 2012;21:2425–2433. doi: 10.1007/s10531-012-0335-x.

HAWKSWORTH, D.L., LÜCKING, R. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. **Microbiol Spectr.** 2017 Jul;5(4). doi: 10.1128/microbiolspec.FUNK-0052-2016. PMID: 28752818.

HE, C.; KLIONSKY, D. J. Regulation mechanisms and signaling pathways of autophagy. **Annual Review of Genetics**, v. 43, p. 67–93, 2009.

HEITMAN, J.; PRIEST, S. J.; YADAV, V. Advances in understanding the evolution of fungal genome architecture. **F1000Research**, v. 9, 2020.

HENCHOZ, S. et al. Phosphorylation- and ubiquitin-dependent degradation of the cyclin- dependent kinase inhibitor Far1p in budding yeast. **Genes and Development**, v. 11, n. 22, p. 3046–3060, 1997.

HIBBETT, D. S. et al. A higher-level phylogenetic classification of the Fungi. **Mycological Research**, v. 111, n. 5, p. 509–547, 1 maio 2007.

JAMES, T. Y. et al. Toward a Fully Resolved Fungal Tree of Life. **Annual Review of Microbiology**, v. 74, p. 291–313, 2020.

JIANG, C. et al. Mitogen-activated protein kinase signaling in plant pathogenic fungi. **PLoS Pathogens**, v. 14, n. 3, p. 4–11, 2018.

KANEHISA, M. et al. Data, information, knowledge and principle: Back to metabolism in KEGG. **Nucleic Acids Research**, v. 42, n. D1, p. 199–205, 2014.

KANEHISA, M. et al. KEGG as a reference resource for gene and protein annotation. **Nucleic Acids Research**, v. 44, n. D1, p. D457–D462, 2016.

- KANEHISA, M., & GOTO, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. **Nucleic acids research**, 28(1), 27–30. <https://doi.org/10.1093/nar/28.1.27>
- KANEHISA, M., FURUMICHI, M., SATO, Y., ISHIGURO-WATANABE, M., & TANABE, M. (2021). KEGG: integrating viruses and cellular organisms. **Nucleic acids research**, 49(D1), D545–D551. <https://doi.org/10.1093/nar/gkaa970>
- KANEHISA, M., SATO, Y. KEGG Mapper for inferring cellular functions from protein sequences. **Protein Sci.** 2020 Jan;29(1):28-35. doi: 10.1002/pro.3711. Epub 2019 Aug 29. PMID: 31423653; PMCID: PMC6933857.
- KANEHISA, M., SATO, Y., KAWASHIMA, M., FURUMICHI, M., & TANABE, M. (2016). KEGG as a reference resource for gene and protein annotation. **Nucleic acids research**, 44(D1), D457–D462. <https://doi.org/10.1093/nar/gkv1070>
- KASSIR, Y. et al. Transcriptional regulation of meiosis in budding yeast. **International Review of Cytology**, v. 224, p. 111–171, 2003.
- KAWASHIMA, L.; TOSHIKI, S. KEGG API: A Web Service Using SOAP / WSDL. v. 674, p. 673–674, 2003. KHAN, I. A. et al. Multifunction of autophagy-related genes in filamentous fungi. **Microbiological Research**, v. 167, n. 6, p. 339–345, 2012.
- KAWASHIMA, SHUICHI; KATAYAMA, TOSHIKI. KEGG API: A Web Service Using SOAP/WSDL to Access the KEGG System. **Genome Informatics**, v. 14, p. 673–674, 2003.
- KELEMEN, O.; CONVERTINI, P.; ZHANG, Z.; WEN, Y.; SHEN, M.; FALALEEVA, M.; STAMM, S. (2013). Function of alternative splicing. **Gene** 514, 1–30. 10.1016/j.gene.2012.07.083
- KLUTTS, J. S. et al. Glycosyltransferases and their products: cryptococcal variations on fungal themes. **FEMS Yeast Research**, v. 6, n. 4, p. 499–512, jun. 2006.
- KOONIN, E. V. Orthologs, Paralogs, and Evolutionary Genomics. **Annual Review of Genetics**, v. 39, n. 1, p. 309–338, 2005.
- KORMAN, T. P. et al. Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 14, p. 6246–6251, 2010.
- KOSZUL, R.; CABURET, S.; DUJON, B.; FISCHER, G. (2004). Eukaryotic genome evolution through the spontaneous duplication of large chromosomal segments. **The EMBO Journal**, 23 (1), 234–243. <https://doi.org/10.1038/sj.emboj.7600024>
- LE, K. Q.; PRABHAKAR, B. S.; HONG, W. J.; LI, L. C. (2015). Alternative splicing as a biomarker and potential target for drug discovery. **Acta Pharmacol. Sin.** 36, 1212–1218. 10.1038/aps.2015.43
- LEE, J. et al. Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. **Journal of Biological Chemistry**, v. 274, n. 23, p. 16040–16046, 1999.

LI, J. et al. Molecular evolution of the deuterolysin (M35) family genes in coccidioides. **PLoS ONE**, v. 7, n. 2, 2012.

LIANG, J. et al. The ubiquitin-modifying enzyme A20 terminates c-type lectin receptor signals and is a suppressor of host defense against systemic fungal infection. **Infection and Immunity**, v. 88, n. 9, 2020.

LIZICAROVA, I. et al. Human pathogen *Candida dubliniensis*: A cell wall mannan with a high content of β -1,2-linked mannose residues. **Carbohydrate Polymers**, v. 70, n. 1, p. 89–100, ago. 2007.

LIN, C. H. et al. The FUS3 MAPK signaling pathway of the citrus pathogen *Alternaria alternata* functions independently or cooperatively with the fungal redox-responsive AP1 regulator for diverse developmental, physiological and pathogenic processes. **Fungal Genetics and Biology**, v. 47, n. 4, p. 381–391, 2010.

LIU, X. H. et al. Autophagy in plant pathogenic fungi. **Seminars in Cell and Developmental Biology**, v. 57, p. 128–137, 2016.

LIU, X. H. et al. Autophagy vitalizes the pathogenicity of pathogenic fungi. **Autophagy**, v. 8, n. 10, p. 1415–1425, 2012.

LUSSIER, M.; SDICU, A.-M.; BUSSEY, H. The KTR and MNN1 mannosyltransferase families of *Saccharomyces cerevisiae*. **Biochimica et Biophysica Acta (BBA) - General Subjects**, v. 1426, n. 2, p. 323–334, jan. 1999.

MADIGAN, MICHAEL T.; STAHL, DAVID A.; BUCKLEY, DANIEL H.; MARTINKO, JOHN M. **Microbiologia de Brock**. 14. ed. São Paulo: Artmed, 2016. 1032p.

MAEDA, H. et al. A novel non-thermostable deuterolysin from *Aspergillus oryzae*. **Bioscience, Biotechnology and Biochemistry**, v. 80, n. 9, p. 1813–1819, 2016.

MARINHO, G. et al. Potential of the filamentous fungus *Aspergillus niger* AN 400 to degrade Atrazine in wastewaters. **Biocatalysis and Agricultural Biotechnology**, v. 9, n. December 2016, p. 162–167, 2017.

MARTÍNEZ-SOTO, D.; RUIZ-HERRERA, J. Functional analysis of the MAPK pathways in fungi. **Revista Iberoamericana de Micología**, v. 34, n. 4, p. 192–202, 2017.

MAZZA, R. et al. The other side of comparative genomics: Genes with no orthologs between the cow and other mammalian species. **BMC Genomics**, v. 10, p. 1–15, 2009.

MILLE, C. et al. Identification of a new family of genes involved in β -1,2-mannosylation of glycans in *Pichia pastoris* and *Candida albicans*. **Journal of Biological Chemistry**, v. 283, n. 15, p. 9724–9736, 2008.

MOHANTA, T. K.; BAE, H. The diversity of fungal genome. **Biological Procedures Online**, v. 17, n. 1, p. 1–9, 2015.

MOLON, M.; WOZNICKA, O.; ZEBROWSKI, J. Cell wall biosynthesis impairment affects the budding lifespan of the *Saccharomyces cerevisiae* yeast.

Biogerontology, v. 19, n. 1, p. 67–79, fev. 2018.

MOTOYAMA, T. et al. Terpendole E, a kinesin Eg5 Inhibitor, is a key biosynthetic intermediate of indole-diterpenes in the producing fungus *Chaunopycnis alba*.

Chemistry and Biology, v. 19, n. 12, p. 1611–1619, 2012.

NAKATOGAWA, H. et al. Dynamics and diversity in autophagy mechanisms: Lessons from yeast. **Nature Reviews Molecular Cell Biology**, v. 10, n. 7, p. 458–467, 2009.

NARANJO-ORTIZ, M. A.; GABALDÓN, T. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. **Biological Reviews**, v. 94, n. 6, p. 2101–2137, 2019.

NICHOLSON, M. J. et al. Identification of two aflatoxin biosynthesis gene loci in *Aspergillus flavus* and metabolic engineering of *Penicillium paxilli* to elucidate their function. **Applied and Environmental Microbiology**, v. 75, n. 23, p. 7469–7481, 2009.

OGATA, H., GOTO, S., FUJIBUCHI, W., KANEHISA, M. Computation with the KEGG pathway database. **Biosystems**. 1998 Jun-Jul;47(1-2):119-28. doi: 10.1016/s0303-2647(98)00017-3. PMID: 9715755.

ONOE, T. et al. Identification of Two Mannosyltransferases Contributing to Biosynthesis of the Fungal-type Galactomannan α -Core-Mannan Structure in *Aspergillus fumigatus*. **Scientific Reports**, v. 8, n. 1, p. 1–16, 2018.

PALMER, G. E.; ASKEW, D. S.; WILLIAMSON, P. R. The diverse roles of autophagy in medically important fungi. **Autophagy**, v. 4, n. 8, p. 982–988, 2008.

PALUMBO, M.C.; COLOSIMO, A.; GIULIANI, A.; FARINA, L. (2007) Essentiality is an emergent property of metabolic network wiring. **FEBS Lett** 581: 2485–2489.

PATTERSON, M. J. et al. Ybp1 and Gpx3 signaling in *Candida albicans* govern hydrogen peroxide-induced oxidation of the cap1 transcription factor and macrophage escape. **Antioxidants and Redox Signaling**, v. 19, n. 18, p. 2244–2260, 2013.

PERSOON, Christiaan Hendrik (1796). **Observationes Mycologicae** (em latim). Parte 1. Leipzig, (Alemanha): Peter Philipp Wolf. Arquivado do original em 19 de dezembro de 2013.

PERUMAL D, LIM C.S., SAKHARKAR K.R., SAKHARKAR M.K. (2009) 'Load Points' and 'Choke Points' as Nodes for Prioritizing Drug Targets in *Pseudomonas aeruginosa*. **Current Bioinformatics** 4: 48–53.

PESIN, J. A.; ORR-WEAVER, T. L. Regulation of APC/C activators in mitosis and meiosis. **Annual Review of Cell and Developmental Biology**, v. 24, p. 475–499, 2008.

POINTING, S. B. et al. Highly specialized microbial diversity in hyper-arid polar desert. **Proceedings of the National Academy of Sciences**, v. 106, n. 47, p. 19964–19969, 24 nov. 2009.

POLLACK, J. K.; HARRIS, S. D.; MARTEN, M. R. Autophagy in filamentous fungi. **Fungal Genetics and Biology**, v. 46, n. 1, p. 1–8, 2009. QIN, T. et al. GhCyP3 improves the resistance of cotton to *Verticillium dahliae* by inhibiting the E3 ubiquitin ligase activity of GhPUB17. **Plant Molecular Biology**, v. 99, n. 4–5, p. 379–393, 2019.

QIN, T. et al. GhCyP3 improves the resistance of cotton to *Verticillium dahliae* by inhibiting the E3 ubiquitin ligase activity of GhPUB17. **Plant Molecular Biology**, v. 99, n. 4–5, p. 379–393, 22 mar. 2019.

RAHMAN, S.A., SCHOMBURG, D. (2006) Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. **Bioinformatics** 22: 1767–1774.

REGGIORI, F.; KLIONSKY, D. J. Autophagic processes in yeast: Mechanism, machinery and regulation. **Genetics**, v. 194, n. 2, p. 341–361, 2013. RICHARDS, R. **Pro PHP XML and Web Services**. [s.l: s.n.].

RIJHWANI, S. K.; SHANKS, J. V. Effect of subculture cycle on growth and indole alkaloid production by *Catharanthus roseus* hairy root cultures. **Enzyme and Microbial Technology**, v. 22, n. 7, p. 606–611, 1998.

RINEAU, F., ROTH, D., SHAH, F., SMITS, M., JOHANSSON, T., CANBÄCK, B., OLSEN, P. B., PERSSON, P., GRELL, M. N., Lindquist, E., Grigoriev, I. V., Lange, L., & Tunlid, A. (2012). The ectomycorrhizal fungus *Paxillus involutus* converts organic matter in plant litter using a trimmed brown-rot mechanism involving Fenton chemistry. **Environmental microbiology**, 14(6), 1477–1487.
<https://doi.org/10.1111/j.1462-2920.2012.02736.x>

RODRIGUES-POUSADA, C. et al. Yeast AP-1 like transcription factors (Yap) and stress response: A current overview. **Microbial Cell**, v. 6, n. 6, p. 267–285, 2019.

ROGGERO, R. et al. Unraveling the mode of action of the antimalarial choline analog G25 in *Plasmodium falciparum* and *Saccharomyces cerevisiae*. **Antimicrobial Agents and Chemotherapy**, v. 48, n. 8, p. 2816–2824, 2004.

SÁ-CORREIA, I. et al. Drug:H⁺ antiporters in chemical stress response in yeast. **Trends in Microbiology**, v. 17, n. 1, p. 22–31, jan. 2009.

SAIKIA, S. et al. Defining paxilline biosynthesis in *Penicillium paxilli*: Functional characterization of two cytochrome P450 monooxygenases. **Journal of Biological Chemistry**, v. 282, n. 23, p. 16829–16837, 2007.

SAIKIA, S. et al. Four gene products are required for the fungal synthesis of the indole-diterpene, paspaline. **FEBS Letters**, v. 580, n. 6, p. 1625–1630, 2006.

SAIKIA, S. et al. Functional analysis of an indole-diterpene gene cluster for lolitrem B biosynthesis in the grass endosymbiont *Epichloë festucae*. **FEBS Letters**, v. 586, n. 16, p. 2563–2569, 2012.

SAIKIA, S. et al. The genetic basis for indole-diterpene chemical diversity in filamentous fungi. **Mycological Research**, v. 112, n. 2, p. 184–199, 2008.

SAKUNO, E. et al. Aspergillus parasiticus cyclase catalyzes two dehydration steps in aflatoxin biosynthesis. **Applied and Environmental Microbiology**, v. 71, n. 6, p. 2999–3006, 2005.

SHARMA, K. K. Fungal genome sequencing: basic biology to biotechnology. **Critical Reviews in Biotechnology**, v. 36, n. 4, p. 743–759, 2016.

SHERRINGTON, S. L. et al. **Host Sensing by Pathogenic Fungi**. 1. ed. [s.l.] Elsevier Inc., 2018. v. 102

SHIMA, Y. et al. Participation in aflatoxin biosynthesis by a reductase enzyme encoded by vrdA gene outside the aflatoxin gene cluster. **Fungal Genetics and Biology**, v. 46, n. 3, p. 221–231, 2009.

SIEBER, P., VOIGT, K., KÄMMER, P., BRUNKE, S., SCHUSTER, S., & LINDE, J. (2018). Comparative Study on Alternative Splicing in Human Fungal Pathogens Suggests Its Involvement During Host Invasion. **Frontiers in microbiology**, 9, 2313. <https://doi.org/10.3389/fmicb.2018.02313>

SIMAAN, H. et al. The AP-1-like transcription factor ChAP1 balances tolerance and cell death in the response of the maize pathogen Cochliobolus heterostrophus to a plant phenolic. **Current Genetics**, v. 66, n. 1, p. 187–203, 2020.

SIMANIS, V. Events at the end of mitosis in the budding and fission yeasts. **Journal of Cell Science**, v. 116, n. 21, p. 4263–4275, 2003.

SIMPSON D.P. (1979). **Dicionário latino de Cassell** (5 ed.). Londres, Reino Unido: Cassell Ltd. p. 883. ISBN 978-0-304-52257-6.

SINGH, S., MALIK, B. K., & SHARMA, D. K. (2007). Choke point analysis of metabolic pathways in E.histolytica: a computational approach for drug target identification. **Bioinformation**, 2(2), 68–72. <https://doi.org/10.6026/97320630002068>

SINGH, S.; MALIK, B. K.; SHARMA, D. K. Choke point analysis of metabolic pathways in E.histolytica: A computational approach for drug target identification. **Bioinformation**, v. 2, n. 2, p. 68–72, 16 ago. 2007.

SKOWYRA, D. et al. Reconstitution of G , Cyclin and Rbx1. v. 284, n. April, p. 662–665, 1999.

SO, Y.-S. et al. Regulatory Mechanism of the Atypical AP-1-Like Transcription Factor Yap1 in Cryptococcus neoformans. **mSphere**, v. 4, n. 6, p. 1–16, 20 nov. 2019.

SOURIRAJAN, A.; LICHTEN, M. Polo-like kinase Cdc5 drives exit from pachytene during budding yeast meiosis. **Genes and Development**, v. 22, n. 19, p. 2627–2632, 2008.

STAJICH, J. E. Fungal Genomes and Insights into the Evolution of the Kingdom. **The Fungal Kingdom**, v. 5, n. 4, p. 619–633, 2017.

STRULLU-DERRIEN, C. et al. The origin and evolution of mycorrhizal symbioses: from palaeomycology to phylogenomics. **New Phytologist**, v. 220, n. 4, p. 1012–1030, 2018.

TAYLOR CM, WANG Q, ROSA BA, HUANG SCC, POWELL K, et al. (2013) Discovery of Anthelmintic Drug Targets and Drugs Using Chokepoints in Nematode Metabolic Pathways. **PLOS Pathogens** 9(8): e1003505. <https://doi.org/10.1371/journal.ppat.1003505>

TAYLOR, C. M. et al. Discovery of Anthelmintic Drug Targets and Drugs Using Chokepoints in Nematode Metabolic Pathways. **PLoS Pathogens**, v. 9, n. 8, p. e1003505, 1 ago. 2013.

TIMMINS, G. S. & DERETIC, V. Mechanisms of action of isoniazid. **Mol Microbiol** 62, 1220–1227, <https://doi.org/10.1111/j.1365-2958.2006.05467.x> (2006).

TORTORA, Gerard J.; CASE, Christine L.; FUNKE, Berdell R. Microbiologia.12. ed. Artmed, 2016. 964 p.

TUNLID A, TALBOT NJ. Genomics of parasitic and symbiotic fungi. **Curr Opin Microbiol**. 2002 Oct;5(5):513-9. doi: 10.1016/s1369-5274(02)00355-7.

TYAGI, R., ELFAWAL, M.A., WILDMAN, S.A. et al. enzyme. **Sci Rep** 9, 9085 (2019). <https://doi.org/10.1038/s41598-019-45548-7>

UMEDA, T. et al. Molecular basis of fosmidomycin's action on the human malaria parasite Plasmodium falciparum. **Sci Rep** 1, 9, <https://doi.org/10.1038/srep00009> (2011).

VAN DE VEERDONK, F. L.; NETEA, M. G. CBLB ubiquitin ligase: A major regulator of antifungal immunity. **Nature Medicine**, v. 22, n. 8, p. 834–835, 2016.

VÁRNAI, A. et al. Carbohydrate-Binding Modules of Fungal Cellulases. Occurrence in Nature, Function, and Relevance in Industrial Biomass Conversion. In: **Advances in Applied Microbiology**. [s.l.] Academic Press Inc., 2014. v. 88p. 103–165.

VILLARES, A. et al. Lytic polysaccharide monooxygenases disrupt the cellulose fibers structure. **Scientific Reports**, v. 7, n. December 2016, p. 1–9, 2017.

WATKINSON, Sara C; BODDY, Lynne; MONEY, Nicholas P. 2015. **The Fungi**. 3.ed. 2015

WATSON, J. D. et al. **Biologia Molecular do Gene**. Artmed Editora, 2015. Disponível em: <http://books.google.pt/books?id=nHjI5lzBkXcC>

WEBBER, C.; PONTING, C. P. Genes and homology. **Current biology : CB**, v. 14, n. 9, p. 332–333, 2004.

WU, B., HUSSAIN, M., ZHANG, W., STADLER, M., LIU, X., & XIANG, M. (2019). Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. **Mycology**, 10(3), 127–140. <https://doi.org/10.1080/21501203.2019.1614106>

YABE, K.; NAKAJIMA, H. Enzyme reactions and genes in aflatoxin biosynthesis. **Applied Microbiology and Biotechnology**, v. 64, n. 6, p. 745–755, 2004.

YADAV, B. S., RONDA, V., VASHISTA, D. P., & SHARMA, B. (2013). Sequencing and computational approaches to identification and characterization of microbial

organisms. **Biomedical engineering and computational biology**, 5, 43–49.
<https://doi.org/10.4137/BECB.S10886>

YANDELL M, ENCE D. A beginner's guide to eukaryotic genome annotation. **Nat Rev Genet**. 2012 Apr 18;13(5):329-42. doi: 10.1038/nrg3174. PMID: 22510764.

YEH I, HANEKAMP T, TSOKA S, KARP PD, ALTMAN RB (2004) Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. **Genome Res** 14: 917–924.

YU, J. Current understanding on aflatoxin biosynthesis and future perspective in reducing aflatoxin contamination. **Toxins**, v. 4, n. 11, p. 1024–1057, 2012.

ZALUGA, J., STRAGIER, P., BAEYEN, S. et al. Comparative genome analysis of pathogenic and non-pathogenic Clavibacter strains reveals adaptations to their lifestyle. **BMC Genomics** 15, 392 (2014). <https://doi.org/10.1186/1471-2164-15-392>

ZHU, L. LE et al. E3 ubiquitin ligase Cbl-b negatively regulates C-type lectin receptor-mediated antifungal innate immunity. **Journal of Experimental Medicine**, v. 213, n. 8, p. 1555–1570, 2016.

ZHU, X. M. et al. Current opinions on autophagy in pathogenicity of fungi. **Virulence**, v. 10, n. 1, p. 481–489, 2019.

ZHANG, S. Q. et al. Mnn10 Maintains Pathogenicity in Candida albicans by Extending α -1,6-Mannose Backbone to Evade Host Dectin-1 Mediated Antifungal Immunity. **PLOS Pathogens**, v. 12, n. 5, p. e1005617, 4 maio 2016.

YIP, C. L. et al. Cloning and analysis of the Saccharomyces cerevisiae MNN9 and MNN1 genes required for complex glycosylation of secreted proteins. **Proceedings of the National Academy of Sciences**, v. 91, n. 7, p. 2723–2727, 29 mar. 1994.

ID	Nome Científico	Tipo	Grupo Taxonômico	Patógeno	Hospedeiro	Referência Bibliográfica
T05291	<i>Alternaria alternata</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	http://www.scielo.br/pdf/pat/v44n2/v44n2a15.pdf
T03125	<i>Bipolaris oryzae</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	https://www.embrapa.br/busca-de-publicacoes/-/publicacao/984326/relacao-entre-a-mancha-parda-bipolaris-oryzae-na-folha-bandeira-e-a-mancha-de-grao-em-genotipos-de-arroz
T03124	<i>Bipolaris sorokiniana</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	http://www.cnpt.embrapa.br/biblio/do/p_do64_5.htm
T03110	<i>Bipolaris zeicola</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	https://acervodigital.unesp.br/handle/unesp/367393
T01072	<i>Botrytis cinerea</i>	Filamentoso	Ascomycetes; Leotiomyces	Sim	Plantas	https://bsppjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1364-3703.2007.00417.x
T04796	<i>Colletotrichum fioriniae</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://apsjournals.apsnet.org/doi/10.1094/PDIS-09-18-1696-PDN
T02237	<i>Eremothecium cymbalariae</i>	Levedura	Ascomycetes; Saccharomycetes	Sim	Plantas	https://apsjournals.apsnet.org/doi/full/10.1094/PDIS-09-17-1507-PDN
T03103	<i>Eutypa lata</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC92327/
T03132	<i>Fomitiporia mediterranea</i>	Filamentoso	Basidiomycetes	Sim	Plantas	http://www.sipav.org/main/jpp/volumes/0110/011012.pdf
T01038	<i>Fusarium graminearum</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://www.sciencedirect.com/science/article/pii/S1087184507001843
T04882	<i>Fusarium oxysporum</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://bsppjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1364-3703.2009.00538.x
T03456	<i>Fusarium pseudograminearum</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://apsjournals.apsnet.org/doi/10.1094/PDIS-11-17-1840-PDN
T04869	<i>Fusarium verticillioides</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	http://www.scielo.br/pdf/pat/v44n1/v44n1a11.pdf
T03455	<i>Heterobasidion irregulare</i>	Filamentoso	Basidiomycetes	Sim	Plantas	https://www.ncbi.nlm.nih.gov/pubmed/23387115
T03105	<i>Marssonina brunnea</i>	Filamentoso	Ascomycetes; Leotiomyces	Sim	Plantas	https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0329.1984.tb00164.x
T03138	<i>Melampsora larici-populina</i>	Filamentoso	Basidiomycetes	Sim	Plantas	https://www.ncbi.nlm.nih.gov/pubmed/29336199
T01089	<i>Moniliophthora perniciosa</i>	Filamentoso	Basidiomycetes	Sim	Plantas	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1982-56762010000500003
T03137	<i>Moniliophthora roreri</i>	Filamentoso	Basidiomycetes	Sim	Plantas	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-54052015000300183
T01612	<i>Nectria haematococca (Fusarium Solani)</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-41582005000600021
T03126	<i>Neofusicoccum parvum</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	https://ainfo.cnptia.embrapa.br/digital/bitstream/item/46339/1/92-Clisneide.pdf
T01161	<i>Parastagonospora nodorum</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	https://link.springer.com/article/10.1007/s13562-018-0481-7
T03102	<i>Phaeoacremonium minimum</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	http://www.scielo.br/pdf/sp/v43n2/0100-5405-sp-43-2-0124.pdf
T03128	<i>Phanerochaete carnososa</i>	Filamentoso	Basidiomycetes	Sim	Plantas	https://link.springer.com/article/10.1007/s00253-010-2516-4
T03109	<i>Pseudocercospora fijiensis</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005876
T02226	<i>Puccinia graminis</i>	Filamentoso	Basidiomycetes	Sim	Plantas	https://doi.org/10.3389/fpls.2016.00205
T01613	<i>Pyrenophora teres</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	https://doi.org/10.1111/j.1364-3703.2010.00649.x
T01027	<i>Pyricularia oryzae</i> 70-15	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://www.nature.com/articles/s41598-018-36852-9
T01071	<i>Sclerotinia sclerotiorum</i>	Filamentoso	Ascomycetes; Leotiomyces	Sim	Plantas	https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002230
T01021	<i>Ustilago maydis</i>	Dimórfico	Basidiomycetes	Sim	Plantas	https://ainfo.cnptia.embrapa.br/digital/bitstream/item/158285/1/circ-222.pdf
T02227	<i>Verticillium alfalfae</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://link.springer.com/chapter/10.1007/978-3-662-44056-8_4
T03450	<i>Verticillium dahliae</i>	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Plantas	https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/779841/4/ct40.pdf
T02285	<i>Zymoseptoria tritici</i>	Filamentoso	Ascomycetes; Dothideomycetes	Sim	Plantas	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4502449/
T01148	<i>Aspergillus clavatus</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009001200012
T01070	<i>Aspergillus Fischeri</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009001200012
T01090	<i>Aspergillus flavus</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009001200012
T01017	<i>Aspergillus fumigatus</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009001200012
T01016	<i>Aspergillus nidulans</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009001200012
T01030	<i>Aspergillus niger</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009001200012
T01018	<i>Aspergillus oryzae</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009001200012
T00189	<i>Candida albicans</i>	Dimórfico	Ascomycetes; Saccharomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?pid=S0037-86822003000500010&script=sci_arttext
T04850	<i>Candida auris</i>	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?pid=S0037-86822003000500010&script=sci_arttext
T01140	<i>Candida dubliniensis</i>	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?pid=S0037-86822003000500010&script=sci_arttext
T01024	<i>Candida glabrata</i>	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?pid=S0037-86822003000500010&script=sci_arttext
T02488	<i>Candida orthopsilosis</i>	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?pid=S0037-86822003000500010&script=sci_arttext
T01115	<i>Candida tropicalis</i>	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?pid=S0037-86822003000500010&script=sci_arttext
T01149	<i>Clavispora lusitaniae</i> ATCC 42720	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	https://academic.oup.com/femspd/article/23/1/75/491184

T05147	Clavispora lusitaniae CBS 6936	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	https://academic.oup.com/femspd/article/23/1/75/491184
T01114	Coccidioides immitis	Dimórfico	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009000900014
T01332	Coccidioides posadasii	Dimórfico	Ascomycetes; Eurotiomycetes	Sim	Humanos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-37132009000900014
T02232	Cryptococcus gattii	Levedura	Basidiomycetes	Sim	Humanos	https://www.msmanuals.com/pt-br/profissional/doi/10.1007/978-95-001-0000-0_14
T01040	Cryptococcus neoformans var. neoformans B-3501A	Levedura	Basidiomycetes	Sim	Humanos	https://www.msmanuals.com/pt-br/profissional/doi/10.1007/978-95-001-0000-0_14
T00243	Cryptococcus neoformans var. neoformans JEC21	Levedura	Basidiomycetes	Sim	Humanos	https://www.msmanuals.com/pt-br/profissional/doi/10.1007/978-95-001-0000-0_14
T01026	Debaryomyces hansenii (Candida Famata)	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566122/
T02238	Encephalitozoon intestinalis	Microsporidians		Sim	Humanos	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC124803/
T02229	Histoplasma capsulatum	Dimórfico	Ascomycetes; Eurotiomycetes	Sim	Humanos	https://www.msmanuals.com/pt-br/profissional/doi/10.1007/978-95-001-0000-0_14
T01116	Lodderomyces elongisporus	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6141678/
T01043	Malassezia globosa	Levedura	Basidiomycetes	Sim	Humanos	https://www.sciencedirect.com/science/article/pii/S0022202X15526584
T06113	Malassezia restricta	Levedura	Basidiomycetes	Sim	Humanos	https://www.sciencedirect.com/science/article/pii/S0022202X15526584
T01150	Meyerozyma guilliermondii	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	https://www.g3journal.org/content/8/3/755
T03549	Paracoccidioides brasiliensis Pb18	Dimórfico	Ascomycetes; Eurotiomycetes	Sim	Humanos	https://www.msmanuals.com/pt-br/profissional/doi/10.1007/978-95-001-0000-0_14
T02230	Paracoccidioides lutzii Pb01	Dimórfico	Ascomycetes; Eurotiomycetes	Sim	Humanos	https://www.teses.usp.br/teses/disponiveis/5/5133/tde-04082015-090755/publico/MarthaEugeniaUranJimenez.pdf
T01033	Yarrowia lipolytica	Levedura	Ascomycetes; Saccharomycetes	Sim	Humanos	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302869/
T02287	Encephalitozoon hellem	Microsporidians		Sim	Aves e Mamíferos	http://www.fmv.ufrpe.br/revistas/revista_fmv_ufrpe_v19n3_2006/557_558_69_75.pdf
T00105	Encephalitozoon cuniculi	Microsporidians		Sim	Mamíferos	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5090138/
T05555	Sporothrix schenckii	Dimórfico	Ascomycetes; Sordariomycetes	Sim	Mamíferos	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-736X2016000400303
T01611	Trichophyton benhamiae	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Mamíferos	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6290094/
T01615	Trichophyton verrucosum	Filamentoso	Ascomycetes; Eurotiomycetes	Sim	Humanos e Bovinos	https://link.springer.com/article/10.1007/s11046-005-0223-y
T03458	Encephalitozoon romaleae	Microsporidians		Sim	Gafanhoto	https://www.ncbi.nlm.nih.gov/pubmed/20050002
T03104	Metarhizium acridum	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Gafanhoto	https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1001264
T03107	Metarhizium robertsii	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Insetos	https://www.ncbi.nlm.nih.gov/pubmed/22174335
T01614	Nosema ceranae	Microsporidians		Sim	Abelha	https://www.nature.com/articles/s41598-017-14702-4
T05029	Purpureocillium lilacinum	Filamentoso	Ascomycetes; Sordariomycetes	Sim	Nematoides	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-29452018000200403
T01049	Vanderwaltozyma polyspora	Levedura	Ascomycetes; Saccharomycetes	Indef.		Não encontrada
T02487	Kazachstania africana	Levedura	Ascomycetes; Saccharomycetes	Indef.		Não encontrada
T02489	Tetrapisispora blattae (Kluyveromyces blattae)	Levedura	Ascomycetes; Saccharomycetes	Indef.		Não encontrada
T03140	Agaricus bisporus var. bisporus H97	Filamentoso	Basidiomycetes	Não		https://en.wikipedia.org/wiki/Agaricus_bisporus
T03136	Agaricus bisporus var. burnettii JB137-S8	Filamentoso	Basidiomycetes	Não		https://en.wikipedia.org/wiki/Agaricus_bisporus
T03141	Anthracyclops flocculosa (Pseudozyma flocculosa)	Levedura	Basidiomycetes	Não		https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2958.2010.07533.x / Nomenclatura nova: https://link.springer.com/article/10.1007/s11557-015-1114-3
T00167	Ashbya gossypii (Eremothecium gossypii)	Levedura	Ascomycetes; Saccharomycetes	Não		https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5553342/
T03129	Auricularia subglabra	Filamentoso	Basidiomycetes	Não		https://pnwfungi.org/index.php/pnwfungi/article/view/1126
T03106	Baudoinia panamericana	Filamentoso	Ascomycetes; Dothideomycetes	Não		https://search.proquest.com/openview/3985c48bef9e175d83d0b46475d9f8ff/1?pq-origsite=gscholar&cbl=18750&diss=y
T03100	Chaetomium thermophilum	Filamentoso	Ascomycetes; Sordariomycetes	Não		https://www.nature.com/articles/srep20937
T03134	Coniophora puteana	Filamentoso	Basidiomycetes	Não		http://www.scielo.br/pdf/cflo/v19n3/1980-5098-cflo-19-03-00333.pdf
T01275	Coprinopsis cinerea	Filamentoso	Basidiomycetes	Não		https://pt.wikipedia.org/wiki/Coprinopsis_cinerea
T03108	Cordyceps militaris	Filamentoso	Ascomycetes; Sordariomycetes	Não		https://www.sciencedirect.com/science/article/abs/pii/S0367326X10001863
T03133	Dichomitus squalens	Filamentoso	Basidiomycetes	Não		https://www.ncbi.nlm.nih.gov/pubmed/30246402
T04883	Glarea lozoyensis	Filamentoso	Ascomycetes; Leotiomyces	Não		https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3272893/
T03130	Gloeophyllum trabeum	Filamentoso	Basidiomycetes	Não		http://www.scielo.br/scielo.php?pid=S2179-80872013000300015&script=sci_abstract&tlng=pt

T01025	<i>Kluyveromyces lactis</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3425888/
T05269	<i>Kluyveromyces marxianus</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://www.sciencedirect.com/science/article/pii/S1749461310000035
T01092	<i>Komagataella phaffii</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://link.springer.com/article/10.1007/s10295-009-0638-4
T01044	<i>Laccaria bicolor</i>	Filamentoso	Basidiomycetes	Não	https://www.nature.com/articles/nature06556
T01117	<i>Lachancea thermotolerans</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://www.nature.com/articles/s41598-018-33105-7
T02233	<i>Naumovozyma castellii</i> (<i>Saccharomyces castellii</i>)	Levedura	Ascomycetes; Saccharomycetes	Não	https://onlinelibrary.wiley.com/doi/full/10.1002/yea.3218
T02234	<i>Naumovozyma dairenensis</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://onlinelibrary.wiley.com/doi/full/10.1002/yea.3303
T01034	<i>Neurospora crassa</i>	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://www.nature.com/articles/srep05135
T03451	<i>Neurospora tetrasperma</i>	Filamentoso	Ascomycetes; Sordariomycetes	Não	http://repositorio.unicamp.br/bitstream/REPOSIP/255730/1/Costa_ArgentinaSampaio_D.pdf
T04849	<i>Penicillium digitatum</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Não	http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-29452002000200059
T04924	<i>Pestalotiopsis fici</i>	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://pubs.acs.org/doi/abs/10.1021/np900308s
T04865	<i>Phialocephala scopiformis</i>	Filamentoso	Ascomycetes; Leotiomycetes	Não	https://mra.asm.org/content/4/1/e01768-15.short
T06133	<i>Pichia kudriavzevii</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007138
T01048	<i>Podospira anserina</i>	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://www.sciencedirect.com/science/article/abs/pii/S0734975016300672
T01162	<i>Postia placenta</i>	Filamentoso	Basidiomycetes	Não	https://doi.org/10.1111/j.1462-2920.2010.02417.x
T03131	<i>Punctularia strigosozonata</i>	Filamentoso	Basidiomycetes	Não	http://www.funga.fi/Karstenia/Karstenia%2040-1%202000-03.pdf
T00005	<i>Saccharomyces cerevisiae</i> (budding yeast)	Levedura	Ascomycetes; Saccharomycetes	Não	https://www.genetics.org/content/197/1/33
T01023	<i>Scheffersomyces stipitis</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-11-136
T01331	<i>Schizophyllum commune</i>	Filamentoso	Basidiomycetes	Não	https://www.nature.com/articles/nbt.1643
T00076	<i>Schizosaccharomyces pombe</i> (fission yeast)	Levedura	Ascomyc.; Schizosaccharomycetes	Não	https://www.nature.com/articles/nature724
T03135	<i>Serpula lacrymans</i>	Filamentoso	Basidiomycetes	Não	https://www.sciencedirect.com/science/article/pii/S1309104215303329
T01610	<i>Sordaria macrospora</i>	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://www.ncbi.nlm.nih.gov/pubmed/20739093
T03101	<i>Spathaspora passalidarum</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://www.ncbi.nlm.nih.gov/pubmed/22124720
T03127	<i>Stereum hirsutum</i>	Filamentoso	Basidiomycetes	Não	https://www.first-nature.com/fungi/stereum-hirsutum.php
T05270	<i>Sugiyamaella lignohabitans</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://academic.oup.com/femsyr/article/16/4/fow037/2680347
T02235	<i>Tetrapisispora phaffii</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://sfamjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1472-765X.2009.02754.x
T02231	<i>Thermothelomyces thermophilus</i>	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://www.sciencedirect.com/science/article/pii/S0168165619308247
T02228	<i>Thermothielavioides terrestris</i>	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://www.sciencedirect.com/science/article/pii/S0166061619300107
T02236	<i>Torulaspora delbrueckii</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://www.sciencedirect.com/science/article/abs/pii/S0168160509003250
T04795	<i>Trametes versicolor</i>	Filamentoso	Basidiomycetes	Não	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2845472/
T03139	<i>Tremella mesenterica</i>	Filamentoso	Basidiomycetes	Não	https://www.mushroomexpert.com/tremella_mesenterica.html
T02991	<i>Trichoderma reesei</i> QM6a	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://www.nature.com/articles/nbt1403
T06031	<i>Trichoderma reesei</i> RUT C-30	Filamentoso	Ascomycetes; Sordariomycetes	Não	https://www.nature.com/articles/nbt1403
T01282	<i>Tuber melanosporum</i>	Filamentoso	Ascomycetes; Pezizomycetes	Não	https://www.sciencedirect.com/science/article/pii/S0308814610001883
T01132	<i>Uncinocarpus reesii</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Não	https://www.microbiologyresearch.org/content/journal/micro/10.1099/00221287-140-6-1481
T03457	<i>Wallemia ichthyophaga</i>	Filamentoso	Basidiomycetes	Não	https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-617
T03142	<i>Wallemia mellicola</i>	Filamentoso	Basidiomycetes	Não	https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007260
T03099	<i>Yamadazyma tenuis</i> (<i>Candida tenuis</i>)	Levedura	Ascomycetes; Saccharomycetes	Não	https://academic.oup.com/femsyr/article/17/3/fox019/3737663
T01118	<i>Zygosaccharomyces rouxii</i>	Levedura	Ascomycetes; Saccharomycetes	Não	https://www.sciencedirect.com/topics/food-science/zygosaccharomyces
T01091	<i>Penicillium rubens</i>	Filamentoso	Ascomycetes; Eurotiomycetes	Não	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3317369/

KO ID	ANNOTATION	qvalue	bootstrap(0.05)	F	NF
K19356	E1.14.99.54; lytic cellulose monooxygenase (C1-hydroxylating) [EC:1.14.99.54]	5,06E-244	100	241	0
K19564	CTR, HNM1; choline transport protein	3,93E-188	100	186	0
K09043	YAP; AP-1-like transcription factor	5,32E-178	100	176	0
K08197	ARN; MFS transporter, SIT family, siderophore-iron:H ⁺ symporter	1,44E-148	100	147	0
K19305	NPII; deuterolysin [EC:3.4.24.39]	1,65E-143	100	142	0
K11232	YPD1; osmolarity two-component system, phosphorelay intermediate protein YPD1	2,20E-129	100	128	0
K17066	MOX; alcohol oxidase [EC:1.1.3.13]	2,29E-126	100	125	0
K23787	DMA1_2; E3 ubiquitin-protein ligase DMA1/2 [EC:2.3.2.23]	2,28E-124	100	123	0
K13682	BMT, WRY; beta-1,2-mannosyltransferase [EC:2.4.1.-]	2,30E-123	100	122	0
K02143	ATPeFK, ATP19; F-type H ⁺ -transporting ATPase subunit k	2,29E-121	100	120	0
K15564	CTK2; CTD kinase subunit beta	2,26E-119	100	118	0
K17741	GRE2; NADPH-dependent methylglyoxal reductase [EC:1.1.1.283]	2,25E-117	100	116	0
K02307	DAM1; DASH complex subunit DAM1	2,21E-116	100	115	0
K17417	YMR31; small subunit ribosomal protein YMR-31	2,27E-112	100	111	0
K15205	TFC6; transcription factor C subunit 6	2,31E-111	100	110	0
K15267	MTEF1; mitochondrial transcription factor 1	2,36E-110	100	109	0
K18637	AXL2; axial budding pattern protein 2	2,88E-103	100	102	0
K17787	AIM5; altered inheritance of mitochondria protein 5	3,14E-100	100	99	0
K03317	TC.CNT; concentrative nucleoside transporter, CNT family	3,43E-97	100	96	0
K19860	GPA1; guanine nucleotide-binding protein alpha-1 subunit	3,86E-94	100	93	0
K02142	ATPeFJ, ATP18; F-type H ⁺ -transporting ATPase subunit j	4,13E-92	100	91	0
K16230	HAC1; transcriptional activator HAC1	4,13E-92	100	91	0
K14522	POP3; ribonuclease P/MRP protein subunit POP3 [EC:3.1.26.5]	4,28E-91	100	90	0
K17440	MRP49; large subunit ribosomal protein MRP49	4,28E-91	100	90	0
K15204	TFC3; transcription factor C subunit 3	4,40E-90	100	89	0
K17801	MBA1; mitochondrial protein MBA1	4,53E-89	100	88	0
K18107	GAAD; L-glyceraldehyde reductase [EC:1.1.1.372]	4,53E-89	100	88	0
K19717	ATG33; autophagy-related protein 33	4,53E-89	100	88	0
K11552	MAL2, MCM21; central kinetochore subunit Mal2/MCM21	4,65E-88	100	87	0
K22151	SIDF; N5-hydroxy-L-ornithine N5-transacylase	4,65E-88	100	87	0
K11370	AHC1; ADA HAT complex component 1	4,77E-87	100	86	0
K17818	ARD1; D-arabinitol dehydrogenase (NADP+) [EC:1.1.1.287]	4,77E-87	100	86	0
K11377	ELP6; elongator complex protein 6	5,32E-84	100	83	0
K14526	POP7, RPP2; ribonuclease P/MRP protein subunit POP7 [EC:3.1.26.5]	5,32E-84	100	83	0
K09983	K09983; uncharacterized protein	5,49E-83	100	82	0
K11551	MIS15, CHL4; central kinetochore subunit Mis15/CHL4	5,49E-83	100	82	0
K06035	DDI2_3; cyanamide hydratase [EC:4.2.1.69]	6,07E-80	100	79	0
K01264	APE3; aminopeptidase Y [EC:3.4.11.15]	6,27E-79	100	78	0
K14854	RSA3; ribosome assembly protein 3	6,27E-79	100	78	0
K22124	GLIT, ACLT; gliotoxin/aspirochlorine biosynthesis thioredoxin reductase [EC:1.8.1.-]	6,49E-78	100	77	0

K01532	PCA1, CAD2; Cd2+-exporting ATPase [EC:3.6.3.3]	6,75E-77	100	76	0
K08712	ABCG2.SNQ2; ATP-binding cassette, subfamily G (WHITE), member 2, SNQ2	6,75E-77	100	76	0
K22493	ALS; agglutinin-like protein	6,75E-77	100	76	0
K05989	ramA; alpha-L-rhamnosidase [EC:3.2.1.40]	6,98E-76	100	75	0
K14528	POP8; ribonuclease P/MRP protein subunit POP8 [EC:3.1.26.5]	6,98E-76	100	75	0
K15072	SAF1; SCF-associated factor 1	7,28E-75	100	74	0
K15530	RGAE; rhamnogalacturonan acetylesterase [EC:3.1.1.86]	7,83E-73	100	72	0
K18580	RHGA; rhamnogalacturonan hydrolase [EC:3.2.1.171]	7,83E-73	100	72	0
K15532	yteR, yesR; unsaturated rhamnogalacturonyl hydrolase [EC:3.2.1.172]	9,56E-68	100	67	0
K15529	POA1; ADP-ribose 1"-phosphate phosphatase [EC:3.1.3.84]	1,12E-63	100	63	0
K01452	E3.5.1.41; chitin deacetylase [EC:3.5.1.41]	1,32E-59	100	59	0
K18064	ACR2; arsenical-resistance protein 2	1,32E-59	100	59	0
K19357	CELB; cellulase [EC:3.2.1.4]	1,32E-59	100	59	0
K21820	AP01; unspecific peroxygenase [EC:1.11.2.1]	1,37E-58	99	58	0
K22494	K22494; CFEM-containing surface protein	1,43E-57	100	57	0
K09241	GAL4; transcriptional regulatory protein GAL4	1,47E-56	100	56	0
K18578	CEL74A; xyloglucan-specific exo-beta-1,4-glucanase [EC:3.2.1.155]	1,47E-56	100	56	0
K11595	CHP1; chromo domain-containing protein 1	1,53E-55	100	55	0
K11396	YNG2, EAF4; chromatin modification-related protein YNG2	1,58E-54	100	54	0
K01266	dmpA, dap; D-aminopeptidase [EC:3.4.11.19]	1,64E-53	100	53	0
K10438	E1.14.13.63, phacB; 3-hydroxyphenylacetate 6-hydroxylase [EC:1.14.13.63]	1,64E-53	100	53	0
K18103	GAAC, LGA1; L-threo-3-deoxy-hexylosonate aldolase [EC:4.1.2.54]	1,64E-53	100	53	0
K06005	SAP; candidapepsin [EC:3.4.23.24]	1,77E-51	100	51	0
K14265	Tam1; tryptophan aminotransferase [EC:2.6.1.27]	1,77E-51	100	51	0
K19844	BEM2; GTPase-activating protein BEM2	1,77E-51	100	51	0
K09248	PUT3; proline utilization trans-activator	1,82E-50	100	50	0
K11696	CNL2; centromere-localized protein 2	1,89E-49	100	49	0
K14770	BUD21, UTP16; U3 small nucleolar RNA-associated protein 16	1,89E-49	100	49	0
K16231	csyA; fungal type III polyketide synthase	1,95E-48	100	48	0
K03785	aroD; 3-dehydroquinate dehydratase I [EC:4.2.1.10]	2,00E-47	100	47	0
K03282	mscL; large conductance mechanosensitive channel	2,08E-46	100	46	0
K14644	OSW3, RRT12; subtilase-type proteinase RRT12 [EC:3.4.21.-]	2,08E-46	100	46	0
K14541	DUR1; urea carboxylase / allophanate hydrolase [EC:6.3.4.6 3.5.1.54]	2,17E-45	100	45	0
K10675	E4.2.1.66; cyanide hydratase [EC:4.2.1.66]	2,23E-44	100	44	0
K13367	FMS1; polyamine oxidase [EC:1.5.3.17]	2,23E-44	100	44	0
K01235	aguA; alpha-glucuronidase [EC:3.2.1.139]	2,30E-43	100	43	0
K19692	TCSA; osmolarity two-component system, sensor histidine kinase TcsA [EC:2.7.13.3]	2,30E-43	100	43	0
K09424	REB1; Myb-like DNA-binding protein REB1	2,38E-42	100	42	0
K22993	K22993; bifunctional Delta-12/omega-3 fatty acid desaturase [EC:1.14.19.6 1.14.19.-]	2,38E-42	100	42	0
K00004	BDH, butB; (R,R)-butanediol dehydrogenase / meso-butanediol dehydrogenase / diacetyl reductase [EC:1.1.1.4 1.1.1.- 1.1.1.303]	2,46E-41	100	41	0

K09102	INO; heteromeric Ino2p/Ino4p transcription factor	2,46E-41	100	41	0
K12411	YOX1, YHP1; homeobox protein YOX1/YHP1	2,55E-40	100	40	0
K20832	EN3GAL; galactan endo-beta-1,3-galactanase [EC:3.2.1.181]	2,55E-40	100	40	0
K00018	hprA; glycerate dehydrogenase [EC:1.1.1.29]	2,63E-39	100	39	0
K14655	RIB2, PUS8; tRNA pseudouridine32 synthase / 2,5-diamino-6-(5-phospho-D-ribitylamino)-pyrimidin-4(3H)-one deaminase [EC:5.4.99.28]	2,71E-38	100	38	0
K20060	ECM21, CSR2, ART2_8; arrestin-related trafficking adapter 2/8	2,71E-38	100	38	0
K22032	K22032; lytic cellulose monooxygenase (C4-dehydrogenating) [EC:1.14.99.56]	2,71E-38	100	38	0
K09423	BAS1; Myb-like DNA-binding protein BAS1	2,80E-37	100	37	0
K17665	CBS2; cytochrome b translational activator protein CBS2, mitochondrial	2,80E-37	100	37	0
K17891	ERV2; FAD-linked sulfhydryl oxidase [EC:1.8.3.2]	2,80E-37	100	37	0
K06659	CLB3_4; G2/mitotic-specific cyclin 3/4	2,90E-36	100	36	0
K14642	YND1; golgi apyrase [EC:3.6.1.5]	2,90E-36	100	36	0
K19846	SPA2; protein SPA2	2,90E-36	100	36	0
K22118	GLIP, ACLP; gliotoxin/aspirochlorine biosynthesis peptide synthetase [EC:6.3.2.-]	2,90E-36	100	36	0
K06912	tfdA; alpha-ketoglutarate-dependent 2,4-dichlorophenoxyacetate dioxygenase [EC:1.14.11.-]	2,99E-35	100	35	0
K09242	LEU3; transcriptional regulatory protein LEU3	2,99E-35	100	35	0
K10245	ELO2; fatty acid elongase 2 [EC:2.3.1.199]	2,99E-35	100	35	0
K10731	SLD3; DNA replication regulator SLD3	2,99E-35	100	35	0
K11762	RSC8, SWH3, SSR2; chromatin structure-remodeling complex subunit RSC8	2,99E-35	100	35	0
K17415	MRP21; small subunit ribosomal protein MRP21	2,99E-35	100	35	0
K22483	HMO1; transcriptional regulator HMO1	2,99E-35	100	35	0
K00493	E1.14.14.1; unspecific monooxygenase [EC:1.14.14.1]	3,06E-34	100	34	0
K01405	PRD1; saccharolysin [EC:3.4.24.37]	3,06E-34	100	34	0
K03034	RPN4; 26S proteasome regulatory subunit N4	3,06E-34	100	34	0
K05534	HOC1; mannan polymerase II complex HOC1 subunit [EC:2.4.1.-]	3,06E-34	100	34	0
K06657	PCL2; G1/S-specific cyclin PLC2	3,06E-34	100	34	0
K06681	LTE1; Gdp/GTP exchange factor required for growth at low temperatures	3,06E-34	100	34	0
K14828	RIX1, IPI2; pre-rRNA-processing protein RIX1	3,06E-34	100	34	0
K15457	SAP155; SIT4-associating protein SAP155	3,06E-34	100	34	0
K18561	FRD; FAD-dependent fumarate reductase [EC:1.3.8.-]	3,06E-34	100	34	0
K18737	SHE3; SWI5-dependent HO expression protein 3	3,06E-34	100	34	0
K22496	RFG1, ROX1; repressor RFG1/ROX1	3,06E-34	100	34	0
K01440	PNC1; nicotinamidase [EC:3.5.1.19]	3,10E-33	100	33	0
K06009	YPS1_2; yapsin 1/2 [EC:3.4.23.41]	3,10E-33	100	33	0
K06661	RAD9; DNA repair protein RAD9	3,10E-33	100	33	0
K09464	GCN4; general control protein GCN4	3,10E-33	100	33	0
K10992	SFR1; Swi5-dependent recombination DNA repair protein 1	3,10E-33	100	33	0
K11132	EST1; telomere elongation protein [EC:2.7.7.-]	3,10E-33	100	33	0
K11234	MFAL; mating pheromone alpha-factor	3,10E-33	100	33	0
K11871	UBP7_11; ubiquitin carboxyl-terminal hydrolase 7/11 [EC:3.4.19.12]	3,10E-33	100	33	0
K12732	ARO10; phenylpyruvate decarboxylase [EC:4.1.1.-]	3,10E-33	100	33	0
K15435	CPD1; 2',3'-cyclic-nucleotide 3'-phosphodiesterase [EC:3.1.4.37]	3,10E-33	100	33	0

K18191	FMC1; ATP synthase assembly factor FMC1, mitochondrial	3,10E-33	100	33	0
K22646	SLX8; E3 ubiquitin-protein ligase complex SLX5-SLX8 subunit SLX8 [EC:2.3.2.27]	3,10E-33	100	33	0
K22835	GAS4; 1,3-beta-glucanosyltransferase GAS4 [EC:2.4.1.-]	3,10E-33	100	33	0
K01198	xynB; xylan 1,4-beta-xylosidase [EC:3.2.1.37]	3,12E-32	100	32	0
K06647	MBP1; transcription factor MBP1	3,12E-32	100	32	0
K06649	SWI4; regulatory protein SWI4	3,12E-32	100	32	0
K06665	SSN6, CYC8; general transcriptional corepressor CYC8	3,12E-32	100	32	0
K06684	DBF2; cell cycle protein kinase DBF2 [EC:2.7.11.-]	3,12E-32	100	32	0
K09238	ACE2; metallothionein expression activator	3,12E-32	100	32	0
K09243	UME6; transcriptional regulatory protein UME6	3,12E-32	100	32	0
K11759	RSC4; chromatin structure-remodeling complex subunit RSC4	3,12E-32	100	32	0
K11766	RSC58; chromatin structure-remodeling complex subunit RSC58	3,12E-32	100	32	0
K14023	HRD3; ERAD-associated E3 ubiquitin-protein ligase component HRD3	3,12E-32	100	32	0
K15081	AMN1; antagonist of mitotic exit network protein 1	3,12E-32	100	32	0
K15158	GAL11, MED15; mediator of RNA polymerase II transcription subunit 15, fungi type	3,12E-32	100	32	0
K16946	SHS1; seventh homolog of septin 1	3,12E-32	100	32	0
K17648	ESTA; versiconal hemiacetal acetate esterase [EC:3.1.1.94]	3,12E-32	100	32	0
K17663	CBP6; cytochrome b pre-mRNA-processing protein 6	3,12E-32	100	32	0
K17798	MSS2; mitochondrial protein MSS2	3,12E-32	100	32	0
K17978	NUM1; nuclear migration protein NUM1	3,12E-32	100	32	0
K18658	ASH1; transcriptional regulatory protein ASH1	3,12E-32	100	32	0
K21627	IFH1; protein IFH1	3,12E-32	100	32	0
K21845	YPP1; cargo-transport protein YPP1	3,12E-32	100	32	0
K22121	GLII, ACLI; gliotoxin/aspirochlorine biosynthesis aminotransferase [EC:2.6.1.-]	3,12E-32	100	32	0
K23750	LIP3, AIM22; octanoyl-CoA:protein transferase [EC:2.3.1.-]	3,12E-32	100	32	0
K00309	dmg; dimethylglycine oxidase [EC:1.5.3.10]	3,14E-31	100	31	0
K03003	RPA34; DNA-directed RNA polymerase I subunit RPA34	3,14E-31	100	31	0
K04091	ssuD; alkanesulfonate monooxygenase [EC:1.14.14.5]	3,14E-31	100	31	0
K06646	CLN3; G1/S-specific cyclin CLN3	3,14E-31	100	31	0
K06655	PHO85; negative regulator of the PHO system [EC:2.7.11.22]	3,14E-31	100	31	0
K06667	CAK1; serine/threonine-protein kinase CAK1 [EC:2.7.11.22]	3,14E-31	100	31	0
K11118	RIF1; RAP1-interacting factor 1	3,14E-31	100	31	0
K11401	SAS2; histone acetyltransferase SAS2 [EC:2.3.1.48]	3,14E-31	100	31	0
K11772	SWI3; SWI/SNF complex subunit SWI3	3,14E-31	100	31	0
K11873	UBP10; ubiquitin carboxyl-terminal hydrolase 10 [EC:3.4.19.12]	3,14E-31	100	31	0
K12612	DCP1; mRNA-decapping enzyme subunit 1	3,14E-31	100	31	0
K12782	AMA1; meiosis-specific APC/C activator protein AMA1	3,14E-31	100	31	0
K14533	RPM2; ribonuclease P protein component, mitochondrial [EC:3.1.26.5]	3,14E-31	100	31	0

K14547	UTP8; U3 small nucleolar RNA-associated protein 8	3,14E-31	100	31	0
K14783	CIC1, NSA3; proteasome-interacting protein CIC1	3,14E-31	100	31	0
K14853	RSA1; ribosome assembly protein 1	3,14E-31	100	31	0
K15068	DIA2, POF3; protein DIA2	3,14E-31	100	31	0
K15222	RRN11; RNA polymerase I-specific transcription initiation factor RRN11	3,14E-31	100	31	0
K15903	GON7, PCC2; protein GON7	3,14E-31	100	31	0
K18044	OCA2; tyrosine-protein phosphatase-like protein OCA2	3,14E-31	100	31	0
K18638	BNI4; protein BNI4	3,14E-31	100	31	0
K19782	DSS1, MSU1; exoribonuclease II, mitochondrial [EC:3.1.13.1]	3,14E-31	100	31	0
K21641	RGT1; glucose transport transcription regulator RGT1	3,14E-31	100	31	0
K22009	MCK1; protein kinase MCK1	3,14E-31	100	31	0
K22541	HTD2; 3-hydroxyacyl-thioester dehydratase, fungi type [EC:4.2.1.-]	3,14E-31	100	31	0
K00062	E1.1.1.116; D-arabinose 1-dehydrogenase [EC:1.1.1.116]	3,11E-30	100	30	0
K02544	MEC3; G2-specific checkpoint protein	3,11E-30	100	30	0
K02646	PHO2; regulatory protein PHO2	3,11E-30	100	30	0
K03261	CAF20; cap-associated protein CAF20	3,11E-30	100	30	0
K03396	gfa; S-(hydroxymethyl)glutathione synthase [EC:4.4.1.22]	3,11E-30	100	30	0
K05530	VAN1; mannan polymerase I complex VAN1 subunit [EC:2.4.1.-]	3,11E-30	100	30	0
K06663	DDC1; DNA damage checkpoint protein	3,11E-30	100	30	0
K08787	STE13; dipeptidyl aminopeptidase Ste13 [EC:3.4.14.-]	3,11E-30	100	30	0
K09240	CAT8; transcriptional regulatory protein CAT8	3,11E-30	100	30	0
K09274	K09274; HMG box factor, other	3,11E-30	100	30	0
K09426	RAP1; Myb-like DNA-binding protein RAP1	3,11E-30	100	30	0
K10729	SLD2; DNA replication regulator SLD2	3,11E-30	100	30	0
K10959	RAD59; DNA repair protein RAD59	3,11E-30	100	30	0
K11134	ELG1; telomere length regulation protein	3,11E-30	100	30	0
K11402	SAS4; something about silencing protein 4	3,11E-30	100	30	0
K11682	SWC3; SWR1-complex protein 3	3,11E-30	100	30	0
K11775	SNF12, SWP73; SWI/SNF complex component SWP73	3,11E-30	100	30	0
K12578	LCD1, ddc2; DNA damage checkpoint protein LCD1	3,11E-30	100	30	0
K12771	SPS1; sporulation-specific protein 1 [EC:2.7.11.1]	3,11E-30	100	30	0
K14013	UBX2, SEL1; UBX domain-containing protein 2	3,11E-30	100	30	0
K14017	SNL1; HSP70 co-chaperone SNL1	3,11E-30	100	30	0
K14551	UTP9; U3 small nucleolar RNA-associated protein 9	3,11E-30	100	30	0
K14814	ALB1; ribosome biogenesis protein ALB1	3,11E-30	100	30	0
K14969	SHG1, CPS15; COMPASS component SHG1	3,11E-30	100	30	0
K15138	ROX3, MED19; mediator of RNA polymerase II transcription subunit 19, fungi type	3,11E-30	100	30	0
K17793	TIM12; mitochondrial import inner membrane translocase subunit TIM12	3,11E-30	100	30	0
K17971	BNR1; BNI1-related protein 1	3,11E-30	100	30	0
K18155	AEP3; ATPase expression protein 3, mitochondrial	3,11E-30	100	30	0
K18720	NUP145; nucleoporin NUP145	3,11E-30	100	30	0
K19705	PTC4; protein phosphatase PTC4 [EC:3.1.3.16]	3,11E-30	100	30	0
K19880	CHS6, BCH2; Chs5-Arf1p-binding protein CHS6/BCH2	3,11E-30	100	30	0

K19992	MSO1; protein MSO1	3,11E-30	100	30	0
K20297	VPS51; vacuolar protein sorting-associated protein 51	3,11E-30	100	30	0
K20311	TRS65; trafficking protein particle complex II-specific subunit 65	3,11E-30	100	30	0
K20312	TCA17; TRAPP-associated protein TCA17	3,11E-30	100	30	0
K22645	SLX5; E3 ubiquitin-protein ligase complex SLX5-SLX8 subunit SLX5 [EC:2.3.2.27]	3,11E-30	100	30	0
K01422	AXL1; protease AXL1 [EC:3.4.24.-]	3,13E-29	100	29	0
K02543	MEC1; cell cycle checkpoint protein MEC1	3,13E-29	100	29	0
K03865	GTF1; glutamyl-tRNA(Gln) amidotransferase subunit F [EC:6.3.5.7]	3,13E-29	100	29	0
K06668	GIN4; serine/threonine-protein kinase GIN4 [EC:2.7.11.-]	3,13E-29	100	29	0
K07106	murQ; N-acetylmuramic acid 6-phosphate etherase [EC:4.2.1.126]	3,13E-29	100	29	0
K08293	SMK1; sporulation-specific mitogen-activated protein kinase SMK1 [EC:2.7.11.24]	3,13E-29	100	29	0
K09074	TYE7; fungal twist-like factor	3,13E-29	100	29	0
K09466	ADR1; zinc finger protein ADR1	3,13E-29	100	29	0
K11767	ARP7, SWP61; actin-related protein 7	3,13E-29	100	29	0
K12780	REC8; meiotic recombination protein REC8, fungi type	3,13E-29	100	29	0
K15563	BUR2; protein BUR2	3,13E-29	100	29	0
K17673	ATP22; mitochondrial translation factor ATP22	3,13E-29	100	29	0
K17715	SAM35, TOM38; sorting and assembly machinery component 35	3,13E-29	100	29	0
K18718	NUP60; nucleoporin NUP60	3,13E-29	100	29	0
K19718	BRE5; UBP3-associated protein BRE5	3,13E-29	100	29	0
K19781	HMI1; ATP-dependent DNA helicase HMI1, mitochondrial [EC:3.6.4.12]	3,13E-29	100	29	0
K19843	TUS1; Rho1 guanine nucleotide exchange factor TUS1	3,13E-29	100	29	0
K21155	VAM3; syntaxin VAM3	3,13E-29	100	29	0
K22240	CNL1; biogenesis of lysosome-related organelles complex 1 subunit CNL1	3,13E-29	100	29	0
K00045	E1.1.1.67, mtlK; mannitol 2-dehydrogenase [EC:1.1.1.67]	3,17E-28	100	28	0
K02181	BFA1; cell cycle arrest protein BFA1	3,17E-28	100	28	0
K03508	POLZ2, rev7; DNA polymerase zeta [EC:2.7.7.7]	3,17E-28	100	28	0
K08165	ATR1; MFS transporter, DHA2 family, multidrug resistance protein	3,17E-28	100	28	0
K09468	MSN2_4; zinc finger protein MSN2/4	3,17E-28	100	28	0
K11399	EAF7; chromatin modification-related protein EAF7	3,17E-28	100	28	0
K11482	HOS1; histone deacetylase HOS1 [EC:3.5.1.98]	3,17E-28	100	28	0
K11881	IRC25, POC3; proteasome chaperone 3	3,17E-28	100	28	0
K12776	MEK1; meiosis-specific serine/threonine-protein kinase MEK1 [EC:2.7.11.1]	3,17E-28	100	28	0
K15221	RRN10; RNA polymerase I-specific transcription initiation factor RRN10	3,17E-28	100	28	0
K17667	PET111; protein PET111, mitochondrial	3,17E-28	100	28	0
K17799	PNT1; pentamidine resistance factor, mitochondrial	3,17E-28	100	28	0
K17802	SOM1; mitochondrial inner membrane protease subunit SOM1	3,17E-28	100	28	0
K17980	MDM32; mitochondrial distribution and morphology protein 32	3,17E-28	100	28	0
K18712	HPR1; THO complex subunit HPR1	3,17E-28	100	28	0

K18804	CTM1; [cytochrome c]-lysine N-methyltransferase [EC:2.1.1.59]	3,17E-28	100	28	0
K21156	CCZ1; vacuolar fusion protein CCZ1	3,17E-28	100	28	0
K21293	CYP619; cytochrome P450 family 619	3,17E-28	100	28	0
K21849	KEG1; beta-1,6-glucan synthesis-associated protein KEG1	3,17E-28	100	28	0
K23323	GLIM, ACLM_U; gliotoxin/aspirochlorine biosynthesis O-methyltransferase [EC:2.1.1.-]	3,17E-28	100	28	0
K00063	ARA1; D-arabinose 1-dehydrogenase [EC:1.1.1.117]	3,23E-27	100	27	0
K05821	ARO9; aromatic amino acid aminotransferase II [EC:2.6.1.58 2.6.1.28]	3,23E-27	100	27	0
K06652	FAR1; cyclin-dependent kinase inhibitor FAR1	3,23E-27	100	27	0
K11758	RSC3_30; chromatin structure-remodeling complex subunit RSC3/30	3,23E-27	100	27	0
K14531	SNM1; ribonuclease MRP protein subunit SNM1	3,23E-27	100	27	0
K14543	MTR2; mRNA transport regulator MTR2	3,23E-27	100	27	0
K15140	SRB6, MED22; mediator of RNA polymerase II transcription subunit 22, fungi type	3,23E-27	100	27	0
K17666	PET54; protein PET54, mitochondrial	3,23E-27	100	27	0
K18759	EDC1_2; enhancer of mRNA-decapping protein 1/2	3,23E-27	100	27	0
K19823	NAO; nitroalkane oxidase [EC:1.7.3.1]	3,23E-27	100	27	0
K22725	IPT1; mannosylinositol phosphoceramide inositolphosphotransferase [EC:2.7.1.-]	3,23E-27	100	27	0
K01083	E3.1.3.8; 3-phytase [EC:3.1.3.8]	3,27E-26	100	26	0
K09245	UGA3; transcriptional activator protein UGA3	3,27E-26	100	26	0
K11133	EST3; telomere replication protein	3,27E-26	100	26	0
K13357	SPO1; putative meiotic phospholipase SPO1 [EC:3.1.1.-]	3,27E-26	100	26	0
K15079	SLX4; structure-specific endonuclease subunit SLX4	3,27E-26	100	26	0
K15132	MED11; mediator of RNA polymerase II transcription subunit 11, fungi type	3,27E-26	100	26	0
K15908	EAP1; protein EAP1	3,27E-26	100	26	0
K17672	AEP2; ATPase expression protein 2, mitochondrial	3,27E-26	100	26	0
K18717	ASM4, NUP59; nucleoporin ASM4	3,27E-26	100	26	0
K19836	HPT1; hypoxanthine phosphoribosyltransferase [EC:2.4.2.8]	3,27E-26	100	26	0
K09235	RME1; zinc finger protein RME1	3,36E-25	100	25	0
K11239	STE5; pheromone-response scaffold protein	3,36E-25	100	25	0
K11879	PBA1, POC1; proteasome chaperone 1	3,36E-25	100	25	0
K12772	SPS4; sporulation-specific protein 4	3,36E-25	100	25	0
K14263	BNA7; kynurenine formamidase [EC:3.5.1.9]	3,36E-25	100	25	0
K14804	SLX9; ribosome biogenesis protein SLX9	3,36E-25	100	25	0
K03379	chnB; cyclohexanone monooxygenase [EC:1.14.13.22]	3,45E-24	100	24	0
K12778	HOP1; meiosis-specific protein HOP1	3,45E-24	100	24	0
K12814	PRP2; pre-mRNA-splicing factor ATP-dependent RNA helicase-like protein PRP2 [EC:3.6.4.13]	3,45E-24	100	24	0
K15217	RRN5; RNA polymerase I-specific transcription initiation factor RRN5	3,45E-24	100	24	0
K21830	pchP; phosphorylcholine phosphatase	3,45E-24	100	24	0
K22241	SNN1; biogenesis of lysosome-related organelles complex 1 subunit SNN1	3,45E-24	100	24	0
K22833	GAS2; 1,3-beta-glucanosyltransferase GAS2 [EC:2.4.1.-]	3,45E-24	100	24	0
K01218	gmuG; mannan endo-1,4-beta-mannosidase [EC:3.2.1.78]	3,54E-23	100	23	0

K11245	MID2, MTL1; mating pheromone-induced death protein 2	3,54E-23	100	23	0
K18852	HPA3; D-amino-acid N-acetyltransferase [EC:2.3.1.36]	3,54E-23	100	23	0
K17738	ARD; D-arabinitol 2-dehydrogenase [EC:1.1.1.250]	3,65E-22	100	22	0
K18786	E2.4.1.321; cellobionic acid phosphorylase [EC:2.4.1.321]	3,65E-22	100	22	0
K06648	SWI6; regulatory protein SWI6	3,70E-21	100	21	0
K11235	MFA; mating pheromone A-factor	3,70E-21	100	21	0
K15265	PUS5; 21S rRNA pseudouridine2819 synthase [EC:5.4.99.43]	3,70E-21	100	21	0
K12851	PRP38; pre-mRNA-splicing factor 38	3,72E-20	100	20	0
K00730	OST4; oligosaccharyl transferase complex subunit OST4	3,69E-19	100	19	0
K05538	MNN1; alpha 1,3-mannosyltransferase [EC:2.4.1.-]	3,69E-19	100	19	0
K06650	CLN2; G1/S-specific cyclin CLN2	3,69E-19	100	19	0
K10868	XRS2; DNA repair protein XRS2	3,69E-19	100	19	0
K11582	SGO1; shugoshin	3,69E-19	100	19	0
K01111	E3.1.3.68; 2-deoxyglucose-6-phosphatase [EC:3.1.3.68]	3,59E-18	100	18	0
K06656	PCL1; G1/S-specific cyclin PLC1	3,59E-18	100	18	0
K11117	STN1; protein STN1	3,59E-18	100	18	0
K12575	SLK19; kinetochore protein SLK19	3,59E-18	100	18	0
K12773	SPR3; sporulation-regulated protein 3	3,59E-18	100	18	0
K15320	ATX, chlB1, mdpB; 6-methylsalicylic acid synthase [EC:2.3.1.165]	3,59E-18	100	18	0
K17674	SOV1; protein SOV1, mitochondrial	3,59E-18	100	18	0
K18736	SHE2; SWI5-dependent HO expression protein 2	3,59E-18	100	18	0
K21294	GLIC, ACLC; gliotoxin/aspirochlorine biosynthesis cytochrome P450 monooxygenase [EC:1.-.-.]	3,59E-18	100	18	0
K01091	gph; phosphoglycolate phosphatase [EC:3.1.3.18]	3,46E-17	100	17	0
K01133	betC; choline-sulfatase [EC:3.1.6.6]	3,46E-17	100	17	0
K06045	shc; squalene-hopene/tetraprenyl-beta-curcumen cyclase [EC:5.4.99.17 4.2.1.129]	3,46E-17	100	17	0
K09247	CYP1, HAP1; heme activator protein 1	3,46E-17	100	17	0
K11554	CTF19; central kinetochore subunit CTF19	3,46E-17	100	17	0
K12853	SNU114; 114 kDa U5 small nuclear ribonucleoprotein component	3,46E-17	100	17	0
K14024	USA1; U1 SNP1-associating protein 1	3,46E-17	100	17	0
K15453	PUS6; tRNA pseudouridine31 synthase [EC:5.4.99.42]	3,46E-17	100	17	0
K22062	PRO1; Delta6-protoilludene synthase [EC:4.2.3.135]	3,46E-17	100	17	0
K00155	LNAA, LNBA; L-tyrosine reductase [EC:1.2.1.101]	3,36E-16	100	16	0
K03001	RPA14; DNA-directed RNA polymerase I subunit RPA14	3,36E-16	100	16	0
K06651	CLB5_6; S-phase entry cyclin 5/6	3,36E-16	100	16	0
K08340	ATG16; autophagy-related protein 16	3,36E-16	100	16	0
K09244	AMDR; transcriptional regulatory protein AMDR	3,36E-16	100	16	0
K11243	FUS1; nuclear fusion protein	3,36E-16	100	16	0
K11573	SPC34; DASH complex subunit SPC34	3,36E-16	100	16	0
K11575	CEP3, CBF3B; centromere DNA-binding protein complex CBF3 subunit B	3,36E-16	100	16	0
K11697	FTA7, CNL3; kinetochore protein Fta7	3,36E-16	100	16	0
K12577	FOB1; DNA replication fork-blocking protein FOB1	3,36E-16	100	16	0
K15207	TFC8; transcription factor C subunit 8	3,36E-16	100	16	0

K17863	PPOA; linoleate 8R-lipoxygenase / 9,12-octadecadienoate 8-hydroperoxide 8R-isomerase [EC:1.13.11.60 5.4.4.5]	3,36E-16	100	16	0
K22119	GLIG, ACLG; gliotoxin/aspirochlorine biosynthesis glutathione S-transferase [EC:2.5.1.-]	3,36E-16	100	16	0
K23006	FAD3; omega-3 fatty acid desaturase (Delta-15 desaturase) [EC:1.14.19.-]	3,36E-16	100	16	0
K03085	SIC1; substrate and inhibitor of the cyclin-dependent protein kinase CDC28	3,22E-15	100	15	0
K06673	SCC4; cohesin loading factor subunit SCC4	3,22E-15	100	15	0
K06687	NET1, ESC5; nucleolar protein involved in exit from mitosis	3,22E-15	100	15	0
K08335	ATG14; autophagy-related protein 14	3,22E-15	100	15	0
K08338	ATG10; ubiquitin-like-conjugating enzyme ATG10	3,22E-15	100	15	0
K10957	RAD55; DNA repair protein RAD55	3,22E-15	100	15	0
K10991	SAE3, SWI5; DNA repair protein Swi5/Sae3	3,22E-15	100	15	0
K11115	CDC13; cell division control protein 13	3,22E-15	100	15	0
K11371	AHC2; ADA HAT complex component 2	3,22E-15	100	15	0
K11555	OKP1; central kinetochore subunit OKP1	3,22E-15	100	15	0
K11556	AME1; central kinetochore subunit AME1	3,22E-15	100	15	0
K11570	DUO1; DASH complex subunit DUO1	3,22E-15	100	15	0
K11574	CBF2, CBF3A, CTF14; centromere DNA-binding protein complex CBF3 subunit A	3,22E-15	100	15	0
K11576	CTF13, CBF3C; centromere DNA-binding protein complex CBF3 subunit C	3,22E-15	100	15	0
K11677	IES3; Ino eighty subunit 3	3,22E-15	100	15	0
K11679	IES5; Ino eighty subunit 5	3,22E-15	100	15	0
K11683	SWC7; SWR1-complex protein 7	3,22E-15	100	15	0
K11764	LDB7, RSC14; chromatin structure-remodeling complex protein RSC14	3,22E-15	100	15	0
K11776	SWP82; SWI/SNF complex component SWP82	3,22E-15	100	15	0
K12414	TAH11; DNA replication licensing factor, required for pre-replication complex assembly	3,22E-15	100	15	0
K12766	MDS1, RIM11; serine/threonine-protein kinase MDS1/RIM11 [EC:2.7.11.1]	3,22E-15	100	15	0
K12770	GIP1; GLC7-interacting protein 1	3,22E-15	100	15	0
K12777	RED1; protein RED1	3,22E-15	100	15	0
K12779	SUM1; suppressor of MAR1-1 protein	3,22E-15	100	15	0
K15080	SAW1; single-strand annealing weakened protein 1	3,22E-15	100	15	0
K15145	MED1; mediator of RNA polymerase II transcription subunit 1, fungi type	3,22E-15	100	15	0
K15150	CSE2, MED9; mediator of RNA polymerase II transcription subunit 9, fungi type	3,22E-15	100	15	0
K15154	MED2; mediator of RNA polymerase II transcription subunit 2	3,22E-15	100	15	0
K15155	MED3, HRS1, PGD1; mediator of RNA polymerase II transcription subunit 3	3,22E-15	100	15	0
K15220	RRN9; RNA polymerase I-specific transcription initiation factor RRN9	3,22E-15	100	15	0
K15543	REF2; RNA end formation protein 2	3,22E-15	100	15	0
K16314	ELM1; serine/threonine-protein kinase ELM1 [EC:2.7.11.1]	3,22E-15	100	15	0
K17670	PET494; COX3 mRNA-specific translational activator PET494	3,22E-15	100	15	0
K17789	AIM37; altered inheritance of mitochondria protein 37	3,22E-15	100	15	0
K17909	ATG23; autophagy-related protein 23	3,22E-15	100	15	0
K17977	YPT11; GTP-binding protein YPT11	3,22E-15	100	15	0
K18713	THP2; THO complex subunit THP2	3,22E-15	100	15	0

K18714	MFT1; THO complex subunit MFT1	3,22E-15	100	15	0
K19830	YCK3; casein kinase I homolog 3 [EC:2.7.11.1]	3,22E-15	100	15	0
K19849	MSB2; signaling mucin MSB2	3,22E-15	100	15	0
K19898	SAG1, AGAL1; alpha-agglutinin	3,22E-15	100	15	0
K20613	KHS; kievitone hydratase [EC:4.2.1.95]	3,22E-15	100	15	0
K21145	ATG38; autophagy-related protein 38	3,22E-15	100	15	0
K21292	CYP503; ent-kaurene oxidase [EC:1.14.14.86]	3,22E-15	100	15	0
K21628	MSA1; G1-specific transcription factors activator MSA1	3,22E-15	100	15	0
K21646	GCR1; glycolytic genes transcriptional activator GCR1	3,22E-15	100	15	0
K21671	GCR2; glycolytic genes transcriptional activator GCR2	3,22E-15	100	15	0
K22239	BLS1; biogenesis of lysosome-related organelles complex 1 subunit BLS1	3,22E-15	100	15	0
K22260	PEX15; peroxin-15	3,22E-15	100	15	0
K01233	csn; chitosanase [EC:3.2.1.132]	3,15E-14	100	14	0
K02627	PDS1; cell cycle regulator	3,15E-14	100	14	0
K05536	MNN5; alpha 1,2-mannosyltransferase [EC:2.4.1.-]	3,15E-14	100	14	0
K08075	E3.1.3.32; polynucleotide 3'-phosphatase [EC:3.1.3.32]	3,15E-14	100	14	0
K09384	K09384; uncharacterized protein	3,15E-14	100	14	0
K11379	YNG1; protein YNG1	3,15E-14	100	14	0
K11398	EAF5; chromatin modification-related protein EAF5	3,15E-14	100	14	0
K11571	HSK3; DASH complex subunit HSK3	3,15E-14	100	14	0
K11774	SNF11; SWI/SNF complex component SNF11	3,15E-14	100	14	0
K11882	POC4; proteasome chaperone 4	3,15E-14	100	14	0
K12187	MVB12; ESCRT-I complex subunit MVB12	3,15E-14	100	14	0
K12416	SWM1, APC13; anaphase-promoting complex subunit SWM1	3,15E-14	100	14	0
K12768	ZIP1; synaptonemal complex protein ZIP1	3,15E-14	100	14	0
K14524	POP6; ribonuclease P/MRP protein subunit POP6 [EC:3.1.26.5]	3,15E-14	100	14	0
K15070	MDM30; mitochondrial distribution and morphology protein 30	3,15E-14	100	14	0
K16947	SPR28; sporulation-regulated protein 28	3,15E-14	100	14	0
K17660	CBP1; cytochrome b pre-mRNA-processing protein 1	3,15E-14	100	14	0
K17791	TIM18; mitochondrial import inner membrane translocase subunit TIM18	3,15E-14	100	14	0
K17984	ATG32; autophagy-related protein 32	3,15E-14	100	14	0
K21143	ATG29; autophagy-related protein 29	3,15E-14	100	14	0
K21597	CAS5; cell wall integrity transcriptional regulator CAS5	3,15E-14	100	14	0
K21997	ATG19_34; autophagy-related protein 19/34	3,15E-14	100	14	0
K22122	GLIF; gliotoxin biosynthesis cytochrome P450 monooxygenase [EC:1.-.-.-]	3,15E-14	100	14	0
K22150	SIDH; mevalonyl-CoA hydratase	3,15E-14	100	14	0
K22242	VAB2; biogenesis of lysosome-related organelles complex 1 subunit VAB2	3,15E-14	100	14	0
K03356	APC9; anaphase-promoting complex subunit 9	3,13E-13	100	13	0
K04781	mbtl, irp9, ybtS; salicylate synthetase [EC:5.4.4.2 4.2.99.21]	3,13E-13	100	13	0
K05947	E2.4.1.217; mannosyl-3-phosphoglycerate synthase [EC:2.4.1.217]	3,13E-13	100	13	0
K09249	ABF1, BAF1; ARS-binding factor 1	3,13E-13	100	13	0
K10969	MNT3; alpha 1,3-mannosyltransferase [EC:2.4.1.-]	3,13E-13	100	13	0

K10970	MNN25; alpha 1,2-mannosyltransferase [EC:2.4.1.-]	3,13E-13	100	13	0
K11403	SAS5; something about silencing protein 5	3,13E-13	100	13	0
K11769	HTL1; chromatin structure-remodeling complex protein HTL1	3,13E-13	100	13	0
K12413	WHI5; G1-specific transcriptional repressor WHI5	3,13E-13	100	13	0
K12691	OST5; oligosaccharyl transferase complex subunit OST5	3,13E-13	100	13	0
K12775	LRS4; monopolin complex subunit LRS4	3,13E-13	100	13	0
K15394	ACE1; hybrid polyketide synthase / nonribosomal peptide synthetase ACE1	3,13E-13	100	13	0
K17664	CBS1; cytochrome b translational activator protein CBS1, mitochondrial	3,13E-13	100	13	0
K17671	AEP1; ATPase expression protein 1, mitochondrial	3,13E-13	100	13	0
K17976	MMR1; mitochondrial MYO2 receptor-related protein 1	3,13E-13	100	13	0
K18386	PAXM, ATMM, LTMM; FAD dependent monooxygenase [EC:1.14.13.-]	3,13E-13	100	13	0
K18731	GFD1; mRNA transport factor GFD1	3,13E-13	100	13	0
K19816	MSS11; transcription activator MSS11	3,13E-13	100	13	0
K21144	ATG31; autophagy-related protein 31	3,13E-13	100	13	0
K22120	GLIK, ACLK; gliotoxin/aspirochlorine biosynthesis gamma-glutamylcyclotransferase [EC:4.3.2.-]	3,13E-13	100	13	0
K22238	BLI1; biogenesis of lysosome-related organelles complex 1 subunit BLI1	3,13E-13	100	13	0
K22722	CSG2; inositol phosphorylceramide mannosyltransferase regulatory subunit	3,13E-13	100	13	0
K00664	ATF; alcohol O-acetyltransferase [EC:2.3.1.84]	3,13E-12	100	12	0
K11241	DIG1; down-regulator of invasive growth 1	3,13E-12	100	12	0
K12596	TRF5; poly(A) RNA polymerase	3,13E-12	100	12	0
K12774	MAM1; monopolin complex subunit MAM1	3,13E-12	100	12	0
K12783	SSP1; sporulation-specific protein 1	3,13E-12	100	12	0
K15317	APTA; asperthecin polyketide synthase	3,13E-12	100	12	0
K19799	RPH1; DNA damage-responsive transcriptional repressor / [histone H3]-trimethyl-L-lysine36 demethylase [EC:1.14.11.69]	3,13E-12	100	12	0
K21388	AHR1; adhesion and hyphal regulator 1	3,13E-12	100	12	0
K21633	CZF1; zinc cluster transcription factor CZF1	3,13E-12	100	12	0
K21927	COP3; alpha-murolene/germacrene-A/gamma-murolene synthase [EC:4.2.3.125 4.2.3.23 4.2.3.126]	3,13E-12	100	12	0
K22153	SIDG; fusarinine C N2-transacetylase	3,13E-12	100	12	0
K01383	BAR1; barrierpepsin [EC:3.4.23.35]	3,16E-11	100	11	0
K01706	gudD; glucarate dehydratase [EC:4.2.1.40]	3,16E-11	100	11	0
K01935	bioD; dethiobiotin synthetase [EC:6.3.3.3]	3,16E-11	100	11	0
K03332	fruA; fructan beta-fructosidase [EC:3.2.1.80]	3,16E-11	100	11	0
K06004	E3.4.23.18; aspergillopepsin I [EC:3.4.23.18]	3,16E-11	100	11	0
K10982	LIF1; ligase-interacting factor 1	3,16E-11	100	11	0
K10983	NEJ1, LIF2; non-homologous end-joining protein 1	3,16E-11	100	11	0
K15318	APTb; hydrolase	3,16E-11	100	11	0
K17661	CBP2; cytochrome b pre-mRNA-processing protein 2	3,16E-11	100	11	0
K18280	FTMD; 6-hydroxytryprostatin B O-methyltransferase [EC:2.1.1.293]	3,16E-11	100	11	0
K09465	BRLA; zinc finger protein BrLA	3,18E-10	100	10	0
K14130	FGAPT2; tryptophan 4-dimethylallyltransferase [EC:2.5.1.34]	3,18E-10	100	10	0
K15316	PKSA; noranthrone synthase [EC:2.3.1.221]	3,18E-10	100	10	0

K18387	PAXB, ATMB, LTMB; paspaline synthase	3,18E-10	100	10	0
K18600	LDB18; potein LDB18	3,18E-10	100	10	0
K19690	CHK1; osomolarity two-component system, sensor histidine kinase CHK1 [EC:2.7.13.3]	3,18E-10	100	10	0
K22264	NVJ1; nucleus-vacuole junction protein 1	3,18E-10	100	10	0
K22732	HELE; helvolic acid biosynthesis 3-ketosteroid Delta1-dehydrogenase	3,18E-10	100	10	0
K11002	ERI1; phosphatidylinositol N-acetylglucosaminyltransferase ERI1 subunit	3,23E-09	100	9	0
K11116	TEN1; telomere length regulation protein	3,23E-09	100	9	0
K14858	SRD1; pre-rRNA-processing protein SRD1	3,23E-09	100	9	0
K17746	ChaDH; chanoclavine-I dehydrogenase [EC:1.1.1.1.332]	3,23E-09	100	9	0
K18385	PAXC, ATMC, LTMC; geranylgeranyldiphosphate transferase	3,23E-09	100	9	0
K14279	U5snRNA; U5 spliceosomal RNA	3,27E-08	100	8	0
K21450	ARAR; arabinolytic transcriptional activator AraR	3,27E-08	100	8	0
K22258	ATG36; autophagy-related protein 36	3,27E-08	100	8	0
K14180	ARI1; aristolochene synthase [EC:4.2.3.9]	3,27E-08	99	8	0
K19900	HWP1, RBT1; hyphal wall protein	3,27E-08	99	8	0
K22587	DCI1; Delta3,5-Delta2,4-dienoyl-CoA isomerase [EC:5.3.3.21]	3,27E-08	99	8	0
K23066	SWNH1_2; swainsonine biosynthesis dioxygenase SwnH1/2 [EC:1.14.11.-]	3,27E-08	99	8	0
K01713	pheC; cyclohexadienyl dehydratase [EC:4.2.1.51 4.2.1.91]	3,27E-07	100	7	0
K12249	TRI5; trichodiene synthase [EC:4.2.3.6]	3,27E-07	100	7	0
K14278	U4snRNA; U4 spliceosomal RNA	3,27E-07	100	7	0
K15321	WA; naphtho-gamma-pyrone polyketide synthase	3,27E-07	100	7	0
K18549	ALP1; oryzin [EC:3.4.21.63]	3,27E-07	100	7	0
K18884	FGAMT; 4-dimethylallyltryptophan N-methyltransferase [EC:2.1.1.261]	3,27E-07	100	7	0
K23320	ACLB; aspirochlorine biosynthesis cytochrome P450 monooxygenase [EC:1.-.-.-]	3,27E-07	100	7	0
K14133	CDPNPT; cyclic dipeptide N-dimethylallyltransferase [EC:2.5.1.-]	3,27E-07	99	7	0
K19808	ARG80, ARGR1; arginine metabolism regulation protein I	3,27E-07	99	7	0
K19848	HKR1; signaling mucin HKR1	3,27E-07	99	7	0
K02805	wecE, rffA; dTDP-4-amino-4,6-dideoxygalactose transaminase [EC:2.6.1.59]	3,27E-07	98	7	0
K17649	OMTB, dmtA; demethylsterigmatocystin 6-O-methyltransferase [EC:2.1.1.109]	3,27E-07	98	7	0
K14871	snR40; small nucleolar RNA snR40	3,27E-07	97	7	0
K22245	INUA; inulinase [EC:3.2.1.7]	3,27E-07	97	7	0
K01123	E3.1.4.41; sphingomyelin phosphodiesterase D [EC:3.1.4.41]	3,33E-06	99	6	0
K10852	PENDE; isopenicillin-N N-acyltransferase [EC:2.3.1.164]	3,33E-06	99	6	0
K15319	APTC; monooxygenase	3,33E-06	99	6	0
K17990	VCPO; vanadium chloroperoxidase [EC:1.11.1.10]	3,33E-06	99	6	0
K21899	AUDH; aldose-2-ulose dehydratase [EC:4.2.1.110]	3,33E-06	99	6	0
K01212	sacC, levB; levanase [EC:3.2.1.65]	3,33E-06	98	6	0
K18685	ACOS; ophiobolin F synthase [EC:4.2.3.145]	3,33E-06	97	6	0
K03770	ppiD; peptidyl-prolyl cis-trans isomerase D [EC:5.2.1.8]	3,37E-05	99	5	0
K04835	mal; methylaspartate ammonia-lyase [EC:4.3.1.2]	3,37E-05	99	5	0
K06680	MAD3; spindle assembly checkpoint component MAD3	3,37E-05	98	5	0

K11773	SNF6; SWI/SNF complex component SNF6	3,37E-05	98	5	0
K23321	ACLL; aspirochlorine biosynthesis cytochrome P450 monooxygenase [EC:1.-.-.]	3,37E-05	98	5	0
K00945	cmk; CMP/dCMP kinase [EC:2.7.4.25]	3,37E-05	97	5	0
K01151	nfo; deoxyribonuclease IV [EC:3.1.21.2]	3,37E-05	97	5	0
K11678	IES4; Ino eighty subunit 4	3,37E-05	97	5	0
K15388	APDE; cytochrome P450 monooxygenase	3,37E-05	96	5	0
K21546	GLIZ; C6 transcription factor GliZ	3,37E-05	96	5	0