UNIVERSIDADE DE MOGI DAS CRUZES

Gledison Eric Teixeira

Explorando o panorama mitocondrial de diferentes espécies do gênero *Malassezia*: montagem, anotação e implementação de ferramentas para a análise de mitogenomas

Mogi das Cruzes, SP

2023

UNIVERSIDADE DE MOGI DAS CRUZES

Gledison Eric Teixeira

Explorando o panorama mitocondrial de diferentes espécies do gênero *Malassezia*: montagem, anotação e implementação de ferramentas para a análise de mitogenomas

Dissertação apresentada ao Programa de Pós- Graduação da Universidade de Mogi das Cruzes (UMC) como parte dos requisitos para a obtenção do grau de Mestre em Biotecnologia.

Área de Concentração: Biotecnologia aplicada a Recursos Naturais e Agronegócios (Genômica e Bioinformática).

Orientador: Prof. Dr. Fabiano Bezerra Menegidio

Mogi das Cruzes, SP

2023

FINANCIAMENTO



Ficha catalográfica
Universidade de Mogi das Cruzes - Biblioteca Central

T266e

Teixeira, Gledison Eric
Explorando o panorama mitocondrial de diferentes espécies do gênero Malassezia: montagem, anotação e implementação de ferramentas para a análise de mitogenomas / Gledison Eric Teixeira. -- Mogi das Cruzes, SP: Universidade de Mogi das Cruzes, 2023.
104 f. : il.

Dissertação (Mestrado em Biotecnologia) -- Universidade de Mogi das Cruzes, 2023.
Orientação: Prof. Dr. Fabiano Bezerra Menegidio.

1. Malassezia/genética 2. Genoma mitocondrial 3. Filogenia 4. Biologia computacional 5. Genoma fúngico I. Menegidio, Fabiano Bezerra II. Universidade de Mogi das Cruzes III. Título

CDD 579.5

Elaborada por Maria Cristina Alves Silva Xavier – CRB-8/10077

ATAS

ATA DA SESSÃO PÚBLICA DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO EM BIOTECNOLOGIA DA UNIVERSIDADE DE MOGI DAS CRUZES

Às quatorze horas do dia trinta e um de julho de dois mil e vinte e três, pela Universidade de Mogi das Cruzes, realizou-se a defesa de dissertação intitulada: "EXPLORANDO O PANORAMA MITOCONDRIAL DE DIFERENTES ESPÉCIES DO GÊNERO *MALASSEZIA*: MONTAGEM, ANOTAÇÃO E IMPLEMENTAÇÃO DE FERRAMENTAS PARA A ANÁLISE DE MITOGENOMAS", para obtenção do grau de mestre pelo(a) candidato(a) **Gledison Eric Teixeira**. Tendo sido o número de créditos alcançados pelo(a) mesmo(a) no total de 52 (cinquenta e dois), a saber: 28 unidades de crédito em disciplinas de pós-graduação e 24 unidades de crédito no preparo da dissertação, o(a) aluno(a) perfaz assim os requisitos para obtenção do grau de mestre. A comissão examinadora estava constituída dos senhores professores doutores: Fabiano Bezerra Menegidio, Marcia Aparecida Silva Bissaco, e Daniela Leite Jabes da Universidade de Mogi das Cruzes. Sob a presidência do primeiro, como orientador da dissertação. A sessão pública da defesa da dissertação foi aberta pelo senhor presidente da comissão que apresentou o(a) candidato(a). Em seguida o(a) candidato(a) realizou uma apresentação oral da dissertação. Ao final da apresentação da dissertação, seguiram-se as arguições pelos membros da comissão examinadora. A seguir a comissão, em sessão secreta, conforme julgamento discriminado por cada membro, considerou o(a) candidato(a)

Aprovado (aprovado(a) /reprovado(a)) por Unanimidade (unanimidade/maioria)

Mogi das Cruzes, 31 de julho de 2023.

Comissão Examinadora	Julgamento
Profª Drª Marcia Aparecida Silva Bissaco	Aprovado (aprovado(a) / reprovado(a))
Profª Drª Daniela Leite Jabes	Aprovado (aprovado(a) / reprovado(a))
Prof. Dr. Fabiano Bezerra Menegidio	Aprovado (aprovado(a) / reprovado(a))

gov.br Documento assinado digitalmente FABIANO BEZERRA MENEGIDIO Data: 11/09/2023 08:59:52-0300 Verifique em https://validar.iti.gov.br

DEDICATÓRIA

Dedico este trabalho a Deus, por me conceder saúde e sabedoria para seguir sempre em frente. Por me mostrar que o estudo é a porta para tudo. Obrigada por ser a minha força e omeu guia em todos os momentos. A ti, Senhor, toda honra e toda a glória.

À alma de meu pai, Armenio, e minha eterna senhora mãe, Neide, pelo apoio e incentivo em todos os momentos da minha vida. Por acreditarem em mim, e não medirem esforços para a concretização dos meus sonhos. Sem vocês, nada seria possível. Amo vocês com amor eterno!

Ao meu irmão Luis, você foi meu equilíbrio e porto seguro nos momentos de tribulação.Obrigado por ensinar à nossa família os valores mais preciosos de um ser humano: a humildade, o amor e o respeito ao próximo. Sem você, nada teria acontecido.

Aos meus amigos, irmãos que eu não tive, em especial ao Mestre Alexandre e Dr. David Aciole, são anjos que Deus colocou em meu caminho. Mesmo com a distância, sempre se fizeram presentes na minha vida e estarão sempre em meu coração. Obrigada pelo companheirismo, apoio e amizade incondicional. Amo vocês!

Ao professor Dr. Fabiano B. Menegidio, que me inspirou a ingressar na Biotecnologia.Sempre foi grande exemplo para mim, e hoje, agradeço por ter me incentivado a seguir
no caminho da pós-graduação. Muito obrigada por tudo!

AGRADECIMENTOS

Agradeço sinceramente à Universidade de Mogi das Cruzes, por todos os desafios proporcionados ao longo dessa jornada. Ao corpo docente e à coordenação do Programa de Pós-Graduação em Biotecnologia, sou grato pela contribuição no meu desenvolvimento profissional e pessoal. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), expresso minha gratidão pela concessão da bolsa que tornou possível a realização desta dissertação.

Agradeço a Deus, por sempre colocar pessoas maravilhosas em meu caminho, que me fazem acreditar em um mundo melhor e me encorajam a seguir em frente. Obrigado por me mostrar que o estudo é a porta para todas as oportunidades, e por nunca me abandonar, guiando-me em todos os momentos. Dedico este trabalho à alma do meu pai Armenio e à minha eterna mãe Neide, que nunca mediram esforços para me ensinar o caminho do bem e sempre me apoiaram em todas as etapas da minha vida. Sem vocês, eu não teria chegado até aqui. Meu amor por vocês é incondicional.

À minha família, símbolo de amor e união, agradeço do fundo do coração. A minha eterna namorada Ingrid Brizolla, obrigado por estar presente nos momentos mais difíceis e conturbados, sempre me apoiando com elogios e palavras motivadoras para seguir em frente e não desistir. Acreditarno meu sonho e me incentivar constantemente é um presente inestimável. É reconfortante saber que posso contar com vocês em todos os momentos. Amo vocês!

Ao meu orientador, Professor Dr. Fabiano B. Menegidio, sou imensamente grato pela oportunidade de realizar este trabalho. Agradeço pela confiança e pela paciência em me atender sempre que precisei. Seus ensinamentos foram admiráveis e seus conselhos foram fundamentais para os meus primeiros passos na pós-graduação. Muito obrigado por tudo! Agradeço também ao meu coorientador Dr. David Aciole Barbosa, por toda a ajuda durante a realização deste trabalho. Sua contribuição foi essencial para o desenvolvimento de todas as pesquisas neste Programa de Pós-Graduação. Muito obrigado!

Aos professores do Programa de Pós-Graduação em Biotecnologia, em especial à professora Dra. Daniela L. Jabes, sou grato pelos ensinamentos que vão além dos limites da matéria, sempre prestativa e com uma didática excelente. Agradeço ao Ms. Alexandre pela constante ajuda prontamente oferecida, e ao Ms. René Aduan Jr., por esclarecer todas as dúvidas biológicas. Aos meus queridos amigos de longa data, Francisco Fogaça e Antonio Viviani, e aos meus filhos Gabriel Teixeira e Eric Silas Rangel Teixeira, por existirem e serem a razão e motivação para o meu sucesso. A todos que de alguma forma contribuíram, meus singelos agradecimentos. Muito obrigado!!!

Meus sinceros e eternos agradecimentos!

“A ciência de hoje é a tecnologia de amanhã.”.

— Edward Teller

RESUMO

A diversidade ecológica e funcional dos fungos do gênero *Malassezia* têm despertado interesse crescente em pesquisas biológicas e clínicas. Neste contexto, a mitogenômica emerge como uma poderosa abordagem para explorar a diversidade genômica e evolutiva desses fungos através do estudo de suas mitocôndrias, organelas essenciais para a respiração celular e produção de energia. Nesta dissertação, nos propusemos a investigar a mitogenômica de *Malassezia* por meio de um amplo conjunto de dados e desenvolvimento de ferramentas de bioinformática. Iniciamos nossa investigação realizando uma prospecção de genomas mitocondriais completos disponíveis em bases de dados públicas, como a *Nucleotide* e *Organelle Genome Resources* do NCBI. Essa prospecção nos permitiu obter uma visão abrangente dos mitogenomas de *Malassezia* previamente caracterizados, revelando uma notável diversidade genômica entre as diferentes espécies do gênero. A fim de ampliar ainda mais a diversidade mitogenômica de *Malassezia*, realizamos a prospecção de bibliotecas de sequenciamento de nova geração de genoma completo (*WGS*) de espécies não representadas nos mitogenomas conhecidos. A montagem e anotação desses novos genomas mitocondriais foram conduzidas com sucesso, fornecendo informações detalhadas sobre a estrutura genômica e a organização funcional dessas organelas. Para otimizar a análise dos dados mitogenômicos e possibilitar a reutilização dessas informações por outros pesquisadores, estabelecemos um fluxo de trabalho de bioinformática baseado em trabalhos anteriores realizados no LaBiOmics. Esse fluxo de trabalho permitiu a montagem e anotação eficiente dos genomas mitocondriais de *Malassezia*, fornecendo uma plataforma robusta para investigações futuras. A análise filogenética comparativa dos mitogenomas revelou relações evolutivas interessantes entre as espécies de *Malassezia* e outros fungos, destacando tanto conservações de genes fundamentais quanto variações em elementos genéticos específicos. Essas informações são valiosas para compreender a evolução genômica e a diversidade dentrodo gênero. Para disponibilizar os resultados obtidos e fornecer uma plataforma interativa para pesquisas futuras, implementamos um banco de dados relacional para o armazenamento dos genomas mitocondriais e suas anotações funcionais. Além disso, desenvolvemos uma aplicação web para pesquisas *BLAST* nas sequências mitogenômicas de *Malassezia*, facilitando a identificação rápida de sequências homólogas em diferentes organismos. Por meio dessas análises e ferramentas, construímos um *dataset* curado contendo informações relevantes sobre a genômica estrutural e funcional dos mitogenomas identificados e montados. Esse *dataset* representa uma valiosa fonte de conhecimento para a comunidade científica, permitindo a reutilização dos dados e a realização de novas investigações em diversos contextos científicos e clínicos. Em conclusão, esta dissertação proporcionou uma visão aprofundada da diversidade mitogenômica de *Malassezia* e sua relevância biológica e evolutiva. Os resultados e ferramentas desenvolvidos aqui abrem novos horizontes para pesquisas futuras em mitogenômica de fungos, impulsionando o campo da biologia fúngica e promovendo avanços científicos em prol da compreensão da biodiversidade e das interações complexas entre esses microrganismos e seus hospedeiros. Esperamos que este trabalho inspire novas descobertas e colaborações, contribuindo para o avanço do conhecimento científico e para a saúde humana e ambiental.

Palavras-chave: mitogenômica; *malassezia*; diversidade genômica; bioinformática; filogenética; banco de dados biológico; banco de dados relacional.

ABSTRACT

The ecological and functional diversity of fungi within the genus *Malassezia* has been increasingly attracting attention in biological and clinical research. In this context, mitogenomics emerges as a powerful approach to explore the genomic and evolutionary diversity of these fungi through the study of their mitochondria, essential organelle for cellularrespiration and energy production. In this dissertation, we set out to investigate the mitogenomics of *Malassezia* through a comprehensive dataset and the development of bioinformatics tools. We initiated our investigation by prospecting complete mitochondrial genomes available in public databases, such as NCBI's Nucleotide and Organelle Genome Resources. This prospecting provided us with a comprehensive view of previously characterized *Malassezia* mitogenomes, revealing remarkable genomic diversity among different species within the genus. To further expand the mitogenomic diversity of *Malassezia*,we conducted prospecting of whole-genome sequencing (WGS) libraries of species not represented in the known mitogenomes. The assembly and annotation of these new mitochondrial genomes were successfully carried out, providing detailed information on the genomic structure and functional organization of these organelles. To optimize the analysis of mitogenomic data and enable the reuse of this information by other researchers, we established a bioinformatics workflow based on previous work carried out by LaBiOmics. This workflow allowed for the efficient assembly and annotation of *Malassezia* mitochondrial genomes, providing a robust platform for future investigations. Comparative phylogenetic analysis of themitogenomes revealed interesting evolutionary relationships between *Malassezia* species and other fungi, highlighting both conservation of fundamental genes and variations in specific genetic elements. These insights are valuable for understanding the genomic evolution and diversity within the genus. To make the obtained results available and provide an interactive platform for future research, we implemented a relational database for storing the mitochondrialgenomes and their functional annotations. Additionally, we developed a web application for BLAST searches of *Malassezia* mitogenomic sequences, facilitating rapid identification of homologous sequences in different organisms. Through these analyses and tools, we curated adataset containing relevant information on the structural and functional genomics of the identified and assembled mitogenomes. This dataset represents a valuable source of knowledge for the scientific community, enabling data reuse and the conduct of further investigations in various scientific and clinical contexts. In conclusion, this dissertation provided an in-depth view of the mitogenomic diversity of *Malassezia* and its biological and evolutionary relevance. The results and tools developed here open new horizons for future research in fungal mitogenomics, driving the field of fungalbiology and advancing scientific knowledge towards understanding biodiversity and the complex interactions between these microorganisms and their hosts. We hope that this work inspires new discoveries and collaborations, contributing to scientific advancement and human and environmental health.

Keywords: mitogenomics; *Malassezia*; genomic diversity; bioinformatics; phylogenetics; biological database; relational database.

LISTA DE ILUSTRAÇÕES

Figura 1	*Phylogeny of the mitogenomes of the genus Malassezia*	41
Figura 2	*Interface of the Mitogenome Database*	49
Figura 3	*Search and display interface of the Database*	50
Figura 4	*Genome Browser*	51
Figura 5	*Chado Schema*	52
Figura 6	*Interface of the Malassezia BLAST*	53
Figura 7	*Similarity search result in Malassezia BLAST interface*	54

LISTA DE ABREVIATURAS E SIGLAS

ATP	*Adenosine TriPhosphat*
AIDS	*Síndrome da Imunodeficiência Adquirida*
BED-6	*Browser Extensible Data type 6*
BLAST	*Basic Local Alignment Search Tool*
bd	*Banco de dados*
CDS	*Coding Sequence*
GFF3	*General feature format version 3*
GIY-YIG	*Domínio GIY-YIG*
HEG	*Homing endonuclease*
IgE	*Imunoglobulina E*
ITS	*Internal Transcribed Spacer*
Kb, Kpb	*Kilobase*
LAGLIDADG	*Endonuclease LAGLIDADG*
lncRNAs	*Long Non-Conding RNA*
miRNA	*microRNA*
mtDNA	*DNA mitocondrial*
ncRNA	*Non-Coding RNA*
ORFs	*Open Reading Frames, Fase de leitura aberta*
PCGs	*Protein Coding Genes, Genes Codificadores De Proteínas*
PCR	*Reação em Cadeia da Polimerase*
RNAseP	*Ribonuclease P*
TBL	*Tabular*

SUMÁRIO

1. INTRODUÇÃO 16
2. OBJETIVOS 22
2.1. OBJETIVO GERAL 22
2.2. OBJETIVOS ESPECÍFICOS 22
3. FUNDAMENTAÇÃO TEÓRICA 24
3.1. *MALASSEZIA*: DEFINIÇÃO 24
3.1.1. *MALASSEZIA*: ECOLOGIA 24
3.1.2. *MALASSEZIA*: DOENÇAS ASSOCIADAS 25
3.1.2.1. PITIRÍASE VERSICOLOR (PV) 26
3.1.2.2. DERMATITE SEBORREICA (DS) 26
3.1.2.3. DERMATITE ATÓPICA (DA) 27
3.1.2.4. FOLICULITE POR *MALASSEZIA* 28
3.1.2.5. PAPILOMATOSE CONFLUENTE E RETICULADA 28
3.1.2.6. OUTRAS CONDIÇÕES 29
3.1.3. *MALASSEZIA*: TAXONOMIA 29
4. ARTIGO I: FIVE MITOCHONDRIAL GENOMES OF THE LIPOPHILIC FUNGUS *MALASSEZIA* (*MALASSEZIALES, EXOBASIDIOMYCETES*) 32
5. ARTIGO II: MINING MITOCHONDRIAL GENOMES OF *MALASSEZIA* FUNGI: BUILDING A COMPREHENSIVE DATASET AND DATABASE 43
6. DISCUSSÃO 57
7. CONSIDERAÇÕES FINAIS 60
8. PERSPECTIVAS FUTURAS 61
9. REFERÊNCIAS 63
ANEXOS 74

1. INTRODUÇÃO

Os fungos são um grupo diverso de organismos eucarióticos, que variam desde leveduras unicelulares até microorganismos multicelulares filamentosos e cogumelos, com efeitos benéficos e prejudiciais à saúde humana e ao meio ambiente (Walker, White, 2005; Wu et al., 2019; James et al., 2020; Li et al., 2021). A diversidade fúngica abrange uma ampla variedade de táxons e diferenças associadas em morfologia, ecologia e estrategias de historia de vida (Hyde et al., 2019; Naranjo-Ortiz, Gabaldón, 2020). Existem estimativas que relatama presença de 2 a 5 milhões de espécies de fungos em todo o mundo, pertencentes a 12 filos, a saber: Ascomycota, Basidiomycota, Entorrhizomycota, Chytridiomycota, Monoble- pharidomycota, Neocallimastigomycota, Mucoromycota, Zoop agomycota, Blastocladiomycota, Aphelidiomycota, Cryptomycota / Rozellomycota e Microsporidia (James et al., 2020; Li et al., 2021).O sub-reino *Dikarya* contém aproximadamente 97% de todas as espécies conhecidas,localizadas nos filos Ascomycota, Basidiomycota e Entorrhizomycota, embora apenas 10% dessa diversidade fúngica tenha sido descrita (James et al., 2020; Li et al., 2021). Aproximadamente 70% da diversidade de Dikarya é representada apenas pelo filo *Ascomycota* (Schmidt-Dannert, 2016), sugerindo que a maioria dos grupos de fungos continua sendo subrepresentado. Assim, estudos integrativos para identificar e descrever novas espécies e aqueles que usam filogenias abrangentes e sequenciamento de genomas (tanto nuclear quanto mitocondrial) continuam necessários para melhorar nossa compreensão sobre a evolução, biologia e filogenia dessas espécies extremamente importantes. Assim como outros eucariotos, os fungos possuem mitocôndrias que são responsáveis pela síntese de ATP (trifosfato de adenosina e desempenham diversas funções biológicas nas células fúngicas. São organelas citoplasmáticas bi-membranosas que realizam uma variedade de tarefas, sendo que a produção de energia por meio da fosforilação oxidativa e da síntese de ATP é sua função principal (Kulik et al., 2021). Além disso, as mitocôndrias produzem precursores metabólicos de macromoléculas, como proteínas e lipídios, bem como subprodutos metabólicos, como amônia e espécies reativas de oxigênio. Além disso, essas células desempenham um papel na manutenção da homeostase, na resposta ao estresse e no processo de apoptose (Grahl et al., 2012; Verma et al., 2018; Zardoya, 2020).

Evidências evolutivas sugerem que as mitocôndrias surgiram como resultado da simbiose mais antiga conhecida entre eucariotos e procariotos (Carvalho et al., 2015). A endossimbiose de uma *Alphaproteobacteria* ancestral de vida livre, que foi permanentemente integrada à célula hospedeira, é sugerida como a origem das mitocôndrias (Carvalho et al., 2015; Muñoz-Gómez et al., 2017). As mitocôndrias fúngicas abrigam seu próprio genoma (mitogenoma), com códigos genéticos diferentes do genoma nuclear. O mitogenoma é responsável pelo correto funcionamento das células e pela estabilidade do genoma nuclear. Por outro lado, para manter suas funções e integridade, as mitocôndrias requerem comunicação com o genoma nuclear (Kaniak-Golik, Skoneczna, 2015; Kulik et al., 2021). Nos fungos, a maioria das proteínas mitocondriais é codificada no genoma nuclear, produzida no citosol como precursores e importadas para as mitocôndrias (Bolender et al., 2008). Assim, o *mtDNA* existente contém principalmente genes codificadores de proteínas que são essenciais para o processo de fosforilação oxidativa (Verma et al., 2018). A arquitetura dos mitogenomas pode variar entre os eucariotos e, até o momento,a maior parte do nosso conhecimento sobre a função e disfunção mitocondrial se baseia em estudos realizados em *Saccharomyces cerevisiae*, que demonstram alta tolerância a mutações que desativam a fosforilação oxidativa e a perda de *mtDNA* (Malina et al., 2018). Os mitogenomas fúngicos são geralmente representados como moléculas circulares, embora versões lineares com extremidades teloméricas definidas, e versões existentes como concatêmeros lineares, tenham sido relatadas (Williamson, 2002; Bullerwell, Lang, 2005; Valach et al., 2011; Formaggioni et al., 2021). O proteoma mitocondrial de *S. cerevisiae* inclui cerca de 1.000 proteínas, enquanto 438 foram identificadas em *N. crassa* por meio de estudos proteômicos (Ambrosio et al., 2013). O típico mitogenoma fúngico de *Ascomycota* e *Basidiomycota* (fungos do sub-reino *Dikarya*) contém 14 genes codificadores de proteínas (*atp6*, *8* e *9*; *cob*; *cox1–3*; *nad1–6*, e *4L*), os genes *rnL* e *rnS*, e geralmente de 20 a 31 genes *trn*.O gene *rps3*, que codifica a proteína ribossômica *S3*, também pode ser encontrado em muitos mitogenomas fúngicos, presente nos genomas do sub-reino *Dikarya*, mas não em certos táxons,como o gênero *Candida* (Freel et al., 2015; Formaggioni et al., 2021). Dessa forma, nota-se que os fungos apresentam variação no conteúdo de genes mitocondriais. Por exemplo, os mitogenomas de leveduras (família *Saccharomycetaceae*) carecem de alguns genes (Freel et al., 2015). Além disso, uma distribuição irregular do gene *rnpB*, que codifica a subunidade RNA do RNAse P mitocondrial, também é frequentemente observada. Além disso, em Ascomycota, os genes mitocondriais são geralmente

codificados em uma única fita, mas em ambas as fitas no membros do filo Basidiomycota (Zardoya, 2020).

Os genomas mitocondriais podem variar em tamanho de 12.055 pares de bases (pb) a 500 kilobases (kpb) (James et al., 2013; Liu et al., 2020). *Rozella allomycis*, considerada uma das primeiras linhagens divergentes de fungos, apresenta 12 kb de tamanho de mitogenoma, enquanto o fungo saprotrófico e ectomicorrízico de solo *Morchella crassipes* apresenta 272 kb de tamanho (James et al., 2013; Liu et al., 2020). A variação de tamanho se deve em parte a íntrons, espaçadores intergênicos, inserções deplasmídeos, duplicações parciais *Open Reading Frames* (*ORFs*) sem função definida (*uORFs*) e genes de endonucleases de *homing* (*HEG*), e, em alguns casos, à proliferação de repetições (Hausner, 2003; Liu et al., 2020; Fonseca et al., 2020; Araújo et al., 2021). Além disso, essa modulação do *mtDNA* está diretamente correlacionada com a presença de elementos de DNA acessórios, como íntrons, *HEGs* e *uORFs*(Fonseca et al., 2020; Araújo et al., 2021). Os íntrons dos grupos I e II podem ser encontrados nos mitogenomas fúngicos. A conservação de sequência, a configuração da estrutura secundáriae a forma de hélice são os fatores que determinam a classificação desses íntrons (Hausner et al.,2014). Famílias de *HEG* como *LAGLIDADG* e *GIY-YIG* podem ser codificadas por íntrons do grupo I, que são mais comuns e estão associados à transposição de sequências em várias partes do mitogenoma (Hafez, Hausner, 2012; Kulik et al., 2021). A estrutura, a sequência e o mecanismo de ruptura dos íntrons do tipo I diferem dos íntrons do grupo II. É considerado quea distribuição de íntrons entre diferentes espécies é irregular. Além disso, os mesmos íntrons e seus *HEGs* podem ser encontrados em mitogenomas considerados evolutivamente distantes (Hausner, 2003; Lang et al., 2007; Hausner et al., 2014; Zardoya, 2020). Apresentando uma baixa taxa de recombinação, padrões variados de herança e uma taxa de mutação diferente do genoma nuclear, os mitogenomas podem ser uma ferramenta importante para a pesquisa de relações filogenéticas (Rubinoff, Holland, 2005; Mendoza et al., 2020). Além disso, a amplificação e sequenciamento demitogenomas parciais ou completos são facilitados pela grande quantidade de cópias presentes no citoplasma da célula (Kulik et al.,2021). No entanto, o sequenciamento de mitogenomas de uma variedade maior de espécies fúngicas é necessário para avaliar a presença de um marcador universal para compreender a evolução das principais linhagens fúngicas (Kulik et al., 2015; Smith, 2015; Avin et al., 2017).

Esses padrões de variação no DNA mitocondrial fornecem informações genéticas que são frequentemente suficientemente variáveis para a genômica comparativa, bem como para estudos evolutivos de fungos. O uso de dados genômicos tem contribuído significativamente para o desenvolvimento da taxonomia e das relações filogenéticas dos fungos, levando a revisões taxonômicas e à identificação de novas espécies. No entanto, vale ressaltar que, apesar das datações moleculares, algumas linhagens fúngicas ainda possuem uma posição filogenética incerta (Naranjo-Ortiz, Gabaldón, 2020). A caracterização dos padrões do mitogenoma pode ajudar a esclarecer essas questões. Alguns exemplos recentes de novas perspectivas filogenéticas foram derivados de estudos em fungos nematófagos (Zhang et al., 2020), no cogumelo comestível comercial *Hypsizygus marmoreus* (Wang et al., 2019) e em fungos ectomicorrízicos do gênero *Rhizopogon* (Li et al., 2019). No entanto, vale ressaltar que a estimativa das relações filogenéticas dos fungos com base tanto nos filogenomas mitocondriais quanto nos nucleares pode ser difícil devido às suas trajetórias evolutivas altamente discordantes (De Chiara et al., 2020).

A mitogenômica se apresenta como uma promessa no campo dos diagnósticos fúngicos. A detecção eficiente de fungos é difícil devido à recuperação ineficiente de DNA genômico para PCR, principalmente devido à distribuição muito baixa e desigual de biomassa fúngica nas amostras testadas e à difícil quebra da parede celular quitinosa (Bilska et al., 2018).Marcadores desenvolvidos com base em sequências mitocondriais proporcionam uma detecção altamente sensível de fungos devido à natureza de multi-cópia dos mitogenomas. Também foi observado que o DNA mitocondrial pode ser útil no processo de decifrar a diversidade de espécies crípticas (Kulik et al., 2020). No entanto, o desenvolvimento de novos testes diagnósticos requer sua avaliação em termos de uniformidade e especificidade contra um painel de testes de diferentes fungos-alvo e não alvo. Isso exige uma atualização contínua do conhecimento tanto sobre a taxonomia fúngica quanto sobre a variação do mitogenoma. O investimento contínuo no sequenciamento de genomas nucleares e mitocondriais é necessário para abrir novas possibilidades para diagnósticos aprimorados, que devem estar em conformidade com a classificação taxonômica/natural real dos fungos.

A mitogenômica também pode trazer novas perspectivas sobre os mecanismos de funcionamento das mitocôndrias, patogenicidade fúngica, virulência e resistência a medicamentos. Os fungos ainda pertencem a um grupo de microrganismos em grande parte inexplorados em termos de abordagens proteômicas e genômicas integrativas para

entender osprocessos mitocondriais (Ambrosio et al., 2013; Medina et al., 2020). Hoje, existem muitas questões em aberto em relação ao proteoma mitocondrial e seu papel na modulação dinâmica dessas organelas (Ambrosio et al., 2013; Mendoza et al., 2020). A herança mitocondrial foi explorada apenas em algumas espécies de fungos (*S. cerevisiae*, *Cryptococcus neoformans*, *Microbotryum violaceum* e *Ustilago maydis*) e esses estudos forneceram evidências de uma grande variação nos mecanismos de herança (Mendoza et al., 2020).

Atualmente, estão disponíveis 692 mitogenomas fúngicos no banco de dados *Organelle Genome Resources* (https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/) do *National Center for Biotechnology Information* (NCBI), demonstrando uma discrepância em relação à diversidade de taxa fúngicos representados. Da mesma forma, estão disponíveis aproximadamente 4670 genomas mitocondriais fúngicos no banco de dados *Nucleotide* (https://www.ncbi.nlm.nih.gov/nuccore), também do NCBI. No entanto, mesmo com essa aparente diversidade de recursos genômicos mitocondriais, até o momento, não existe nenhum banco de dados especializado para dados mitogenômicos de fungos, que permitam uma melhorinteração da comunidade científica.

Como os fungos exibem uma grande diversidade de espécies com uma ampla gama defunções ecológicas, selecionamos nesse trabalho o gênero *Malassezia* como modelo para explorar essa diversidade através de um amplo e detalhado conjunto de dados (*dataset*) mitogenômico. Nesse sentido, os objetivos desta dissertação consistiram em: **(1)** prospecção dos genomas mitocondriais montados e completos disponíveis nas bases de dados *Nucleotide* (https://www.ncbi.nlm.nih.gov/nuccore) e/ou *Organelle Genome Resources* (https://www.ncbi.nlm.nih.gov/genome/organelle/) mantidas pelo *National Center for Biotechnology Information* (*NCBI*; https://www.ncbi.nlm.nih.gov/); **(2)** prospecção de bibliotecas de sequenciamento de nova geração de genoma completo (*Whole Genome Sequencing*, *WGS*) de espécies de fungos do gênero *Malassezia* que não possuíssem representatividade entre os mitogenomas observados; **(3)** estabelecimento de um fluxo de trabalho de bioinformática, baseado em trabalhos anteriores realizados pelo Laboratório de Bioinformática e Ciências Ômicas (LaBiOmics), para a montagem e anotação demitogenomas de fungos do gênero *Malassezia*; **(4)** montagem e anotação de novos genomas mitocondriais de fungos do gênero *Malassezia*, com posterior análise filogenética em comparação com outros fungos já disponíveis do mesmo grupo taxonômico; **(5)** construção de um *dataset* curado, organizado e reutilizável, contendo informações relevantes sobre a genômica estrutural e funcional dos mitogenomas

identificados e montados; **(6)** construção e implementação de um banco de dados relacional para o armazenamento dos genomas mitocondriais e anotações funcionais; **(7)** implementação de uma aplicação *web* que permita pesquisas *BLAST* nas sequências mitogenômicas de *Malassezia* presentes no banco de dados.

A justificativa para a realização deste trabalho se baseia na relevância científica e clínica do estudo dos genomas mitocondriais de fungos, principalmente nos patogênicos como os pertencentes ao gênero *Malassezia*. O entendimento da genômica desses fungos pode levar à identificação de alvos terapêuticos, ao desenvolvimento de estratégias de diagnóstico mais precisas e à compreensão dos mecanismos envolvidos na patogenicidade desses microrganismos. Além disso, a implementação de um banco de dados especializado facilitará o acesso e a análise dosgenomas mitocondriais, incentivando a colaboração entre pesquisadorese promovendo avanços na área.

Este trabalho visa fornecer uma contribuição relevante para o campo da genômica de fungos, por meio da implementação de recursos bioinformáticos, como um banco de dados especializado que facilite a análise e o compartilhamento de informações genômicas. Espera- se que esses recursos computacionais se apresentem como uma ferramenta útil para a comunidade científica interessada na compreensão da genômica estrutural e funcional desses fungos.

2. OBJETIVO

2.1. Objetivo Geral

O objetivo geral deste trabalho consiste em explorar a diversidade do gênero *Malassezia*, um grupo de fungos com ampla gama de funções ecológicas, por meio de um extenso e detalhado conjunto de dadosmitogenômicos. Para isso, serão realizadas prospecções de genomas mitocondriais completos disponíveis em bases de dados públicas, bem como bibliotecas de sequenciamento de nova geração de genomas completos de espécies de fungos do gênero *Malassezia* não representadas nos mitogenomas observados. Além disso, será estabelecido um fluxo de trabalho de bioinformática para a montagem e anotação de novos genomas mitocondriais, seguido de análises filogenéticas comparativas com outros fungos do mesmo grupo taxonômico.

2.2. Objetivos Específicos

- Prospecção de Genomas Mitocondriais: realizar a busca e identificação de genomas mitocondriais completos disponíveis nas bases de dados *Nucleotide* e/ou *Organelle Genome Resources* mantidas pelo NCBI;
- Prospecção de Genomas Completos por *WGS*: realizar a busca de bibliotecas de sequenciamento de nova geração de genoma completo (*Whole Genome Sequencing*, *WGS*) de espécies de fungos do gênero *Malassezia* que ainda não possuam representatividade nos mitogenomas observados;
- Montagem e Anotação de Mitogenomas: estabelecer um fluxo de trabalho de bioinformática para a montagem e anotação de mitogenomas de fungos do gênero *Malassezia*, com base em trabalhos anteriores realizados pelo Laboratório de Bioinformática e Ciências Ômicas (LaBiOmics);
- Análise Filogenética Comparativa: realizar análises filogenéticas para comparar os novos genomas mitocondriais de *Malassezia* com outros fungos já disponíveis no mesmo grupo taxonômico, buscando *insights* sobre as relações evolutivas e a diversidade genômica;

- Construção de um *Dataset* Curado: organizar e estruturar um *dataset* curado contendoinformações relevantes sobre a genômica estrutural e funcional dos mitogenomas

  identificados e montados, tornando-o reutilizável para futuras pesquisas;
- Implementação de um Banco de Dados Relacional: desenvolver e implementar umbanco de dados relacional para o armazenamento dos genomas mitocondriais e suas

  anotações funcionais, proporcionando uma plataforma de fácil acesso e consulta aosdados;
- Desenvolvimento de Aplicação Web para Pesquisas *BLAST*: criar uma aplicação web

  que permita pesquisas *BLAST* (*Basic Local Alignment Search Tool*) nas sequências mitogenômicas de *Malassezia*, possibilitando a identificação rápida e precisa de sequências homólogas em diferentes organismos.

Ao atingir esses objetivos, espera-se ampliar o conhecimento sobre a diversidade e evolução do gênero Malassezia, bem como disponibilizar uma valiosa fonte de dados e ferramentas para a comunidade científica envolvida em estudos genômicos e filogenéticos de fungos e áreas correlatas.

3. FUNDAMENTAÇÃO TEÓRICA

3.1. *Malassezia*: definição

Ao longo de mais de um século, as leveduras de *Malassezia* têm sido reconhecidas como organismos lipofílicos, fazendo parte da flora cutânea normal humana, além de serem associadas a certas doenças de pele (Findley et al., 2013; Oh et al., 2014; Oh et al., 2016; Saunteet al., 2020). Além disso, desde a década de 1980, foi relatado que elas podem causar infecções sistêmicas oportunistas (Marcon, Powell, 1992). Essas leveduras têm sido encontradas frequentemente em uma ampla variedade de animais de sangue quente, e sua distribuição mais ampla na natureza está sendo explorada atualmente por meio de técnicas moleculares (Renker et al., 2003; Bjelland, Ekman, 2005; Fell et al., 2006; Gandra et al., 2008; Lai et al.; 2007; Saunte et al., 2020). Características distintivas do gênero *Malassezia* incluem uma morfologia característica e afinidade por lipídios em cultura. Observações microscópicas eletrônicas mostraram que essas leveduras possuem uma parede celular espessa e multicamadas, com brotos formados sucessivamente a partir de um único ponto na célula mãe (Ashbee, Evans, 2002). Isso leva à formação de uma cicatriz proeminente no broto, conferindo a forma característica de uma garrafa tanto à célula mãe quanto ao broto, o que facilita sua identificação em material de pele. Além disso, elas possuem a propriedade fisiológica de serem capazes de utilizar lipídios e, com exceção de *M. pachydermatis*, as espécies apresentam uma necessidade absoluta de lipídios nos meios de cultura. Atualmente, o gênero é classificado na ordem Malasseziales, pertencente aos Exobasidiomycetes (Basidiomycota).

3.1.1 *Malassezia*: ecologia

Nos últimos anos, foram realizadas numerosas pesquisas para investigar a presença das leveduras *Malassezia* na pele humana. Esses estudos utilizaram microscopia direta e técnicas de cultura para identificar e quantificar a colonização dessas leveduras (Theelen et al.,2018; Theelen et al., 2019; Saunte et al., 2020). Os resultados revelaram que as espécies de *Malassezia* estão universalmente presentes na população, com a colonização iniciando na infância e atingindo concentrações máximas durante a puberdade e o início da idade adulta. Aspesquisas se concentraram principalmente no couro cabeludo em relação a condições como dermatite seborreica e caspa, e observou-se que as áreas ricas em glândulas sebáceas apresentam as maiores concentrações dessas leveduras

(Theelen et al., 2018; Theelen et al., 2019). Estudos anteriores identificaram as espécies individuais de *Malassezia* com base no conhecimento disponível na época. No entanto, após a descoberta de novas espécies em 1996 e 2002, além dos avanços nos métodos moleculares para identificação de espécies, estudos subsequentes forneceram informações sobre a distribuição de cada espécie na pele saudável e sua associação com diversas condições dermatológicas (Wikramanayake et al., 2019; Saunte et al., 2020).

Nesse mesmo sentido, as pesquisas em animais têm se concentrado principalmente na recuperação de *Malassezia pachydermatis*, que foi encontrada em uma ampla variedade de hospedeiros animais (Theelen et al., 2018; Theelen et al., 2019). Outras espécies de *Malassezia* também foram identificadas em animais, mas o espectro completo das dezoito espécies e sua distribuição na pele humana e animal, bem como no ambiente não vivo, ainda está sendo investigado (Vijaya Chandra et al., 2021).

3.1.2 *Malassezia*: doenças Associadas

Diferentes espécies de *Malassezia* têm sido associadas a diversas doenças em humanose animais. Isso varia desde infecções "verdadeiras", em que há invasão dos tecidos, como a pitiríase versicolor (PV) (Gordon, 1951; Ashbee et al., 1993; Silva et al., 1997; Nakabayashi et al., 2000; Gupta et al., 2002; Crespo-Erchiga e Florencio, 2006), ou casos mais raros de infecção sistêmica, onde a levedura parece desempenhar um papel patogênico por mecanismos indiretos ou imunológicos, como no caso da dermatite seborreica (DS) (Silva et al., 1997; Sugita et al., 2004; Wananukul et al., 2005; Deangelis et al., 2005; Ro, Dawson, 2005; Kim, 2009; Borda, Wikramanayake, 2015; Honnavar et al., 2016; Islamoglu, 2019; Saunte et al., 2020; Dall'Oglio et al., 2022). Além disso, existem doenças atribuídas à *Malassezia*, mas com evidências científicas insuficientes, como a papilomatose confluente e reticulada - conhecida como síndrome de Gougerot-Carteaud e a onicomicose. Neste último caso, é quase certo que a *Malassezia* seja uma colonizadora do espaço subungueal.

A história das descrições dessas doenças é difícil de ser desvendada, em parte devido ao uso de termos semelhantes para descrever o que agora sabemos serem estados de doença distintos. Por exemplo, Braun-Falco *et al.* utilizaram o termo "pitiríase" para descrever várias condições diferentes, incluindo descamação no couro cabeludo. Em geral, ao longo dosséculos, o termo pitiríase tem sido usado para descrever condições de descamação distintas da psoríaseou da ictiose; a palavra alternativa "*porrigo*" também era frequentemente usada para condiçõessemelhantes de descamação.

3.1.2.1. Pitiríase Versicolor (PV)

A pitiríase versicolor (PV) é uma infecção superficial da pele na qual os organismos, geralmente *Malassezia globosa*, desenvolvem filamentos e causam danos patológicos leves ao estrato córneo da epiderme por meio da penetração nos tecidos e da inflamação resultante (Saunte et al., 2020). A PV é uma das doenças de pele mais comuns em uma ampla gama de áreas geográficas, mas é mais frequente nos trópicos, com maior incidência em crianças mais velhas e adultos jovens (Gupta et al., 2002; Crespo-Erchiga, Florencio, 2006). A PV também é observada regularmente em climas nórdicos, muitas vezes em pacientes que passaram suas férias em um ambiente mais quente. Embora essa infecção possa ocorrer em indivíduos saudáveis, também é relatada em alguns pacientes imunocomprometidos, como aqueles que recebem terapia de longo prazo com corticosteroides sistêmicos e outros medicamentos imunossupressores; no entanto, não é um indicador de infecção por HIV (vírus da imunodeficiência humana) na AIDS (Síndrome da Imunodeficiência Adquirida) (Adalsteinsson et al., 2020).

A PV se apresenta com manchas escamosas hipopigmentadas e hiperpigmentadas, quecom o tempo se tornam confluentes no tronco, cabeça e membros proximais. Pode ser acompanhada de sintomas leves de irritação. Foram descritas formas atípicas que se espalham pelos membros proximais ou produzem alterações atróficas na pele (Gupta et al., 2002; Crespo-Erchiga, Florencio, 2006).

O diagnóstico da infecção é realizado por microscopia direta de escamas da pele; o cultivo de escamas não tem lugar no diagnóstico laboratorial. O tratamentocom uma variedade de antifúngicos, como imidazóis, geralmente é eficaz, embora a aplicação adequada de medicamentos tópicos em áreas extensas da pele seja tecnicamente desafiadora (Gupta et al., 2002). Em pacientes que não residem em regiões tropicais, a recidiva é incomum; no entanto, se os fatores predisponentes não forem removidos, a reinfecção pode ocorrer.

3.1.2.2. Dermatite Seborreica (DS)

A dermatite seborreica (DS) é uma doença de pele escamosa que afeta o couro cabeludo, rosto e tórax (Wikramanayake et al., 2019; Saunte et al., 2020; Dall'Oglio et al., 2022). Há manchas eritematosas escamosas, mais comuns no couro cabeludo, dobras nasolabiais, sobrancelhas, atrás das orelhas e no tórax sobre o esterno. É uma das principais, senão a única, causa de descamação do couro cabeludo ou caspa. O papel patogenético da *Malassezia* na DS ainda é um mistério, pois não há evidências de que os

organismos invadam a pele. No entanto, o entendimento da ligação explícita entre a DS e a *Malassezia* é baseado em evidências observacionais de que a remoção das leveduras da pele, usando um agente antifúngico, geralmente leva à remissão, e a recaída está associada ao reaparecimento dos organismos na superfície da pele (Wikramanayake et al., 2019; Dall'Oglio et al., 2022). Não há um modelo experimental para o estudo daa DS e nem todos os pacientes respondem aos antifúngicos. Dessa forma, apesar das investigações intensivas, os mecanismos patogenéticosenvolvidos ainda são obscuros. Tentativas de demonstrar o envolvimento de viasimunológicas ou bioquímicas específicas não foram bem-sucedidas (Saunte et al., 2020; Dall'Oglio et al., 2022). No entanto, além de afetar pacientes saudáveis, a DS é comum na faseinicial da AIDS e também em pacientes com doenças neurológicas crônicas, como a doença deParkinson. Não foram demonstrados defeitos na função das células T em pacientes com DS queexplicariam a ligação com o HIV (Adalsteinsson et al., 2020). Há algumas evidências de níveismais elevados de anticorpos contra a *Malassezia* em pacientes com DS. No entanto, a base patológica real da DS ainda não está clara. Intuitivamente, o mecanismo pareceria ser a falha em suprimir uma resposta inflamatória a uma levedura comensal superficial, que, como consequência, é ativada em pacientes com DS (Wikramanayake et al., 2019; Dall'Oglio et al., 2022). No entanto, atualmente, essas ideias são especulativas. O diagnóstico ainda é clínico, com base na morfologia clínica característica e distribuição das lesões. O tratamento com shampoos antifúngicos, cremes ou antifúngicos orais geralmente é eficaz, embora haja exceções.

3.1.2.3. Dermatite Atópica (DA)

As espécies de *Malassezia* também têm sido implicadas no desenvolvimento de uma forma de dermatite atópica (DA) que afeta o rosto e o couro cabeludo (Chng et al., 2016; Bjerre et al., 2017; Saunte et al., 2020; Koh et al., 2022). Esse fenômeno é mais comum em adultos jovens do que em crianças e afeta não apenas o rosto, não as dobras cutâneas. A evidência de uma conexão com *Malassezia* baseia-se na observação de que a) os pacientes respondem à terapia antifúngica e b) há uma alta prevalência de anticorpos específicos (IgE) contra antígenosde *Malassezia* nos pacientes afetados (Balaji et al., 2011; Brodska et al., 2014; Glatz et al., 2015; Koh et al., 2022).

O efeito positivo do tratamento é difícil de entender com base em qualquer explicação, exceto a remoção das leveduras da superfície da pele, uma vez que antifúngicos azóis sem atividade contra Staphylococcus aureus, como o cetoconazol, são tão eficazes quanto aqueles com propriedades antibacterianas. Existem precedentes para o

papel da imunidade mediada porIgE no desencadeamento de surtos na DA (Balaji et al., 2011; Brodska et al., 2014; Glatz et al., 2015; Koh et al., 2022), uma vez que um mecanismo semelhante tem sido proposto como contribuinte para os surtos muito mais comuns associadosa *Staphylococcus*.

No último exemplo, há evidências de que outros desencadeadores bacterianos mediados imunologicamente também possam estar envolvidos. O *S. aureus*, por exemplo, é uma fonte de superantígenos que podem causar diretamente a ativação de células T. Portanto, o mecanismo pelo qual a *Malassezia* desempenha um papel no agravamento da DA ainda não é conhecido (Chng et al., 2016; Bjerre et al., 2017; Saunte et al., 2020; Koh et al., 2022).

3.1.2.4. Foliculite por Malassezia

A foliculite por *Malassezia* é uma erupção inflamatória pustular dispersa que é principalmente observada no tronco superior, costas e frente (Abdillah, Ranque, 2021). As pústulas não estão associadas a comedões (cravos) e causam coceira (Saunte *et al.*, 2020). A condição muitas vezes é desencadeada pela exposição ao sol, mas também pode ocorrer em pacientes gravemente enfermos internados no hospital. Também existe uma forma pustular de dermatite seborreica, novamente geralmente no tórax ou nas costas, onde as pústulas são pequenas e rodeadas por eritema. A biópsia da lesão revela um grande número de leveduras dentro dosfolículos capilares e inflamação circundante. Em geral, a terapia oral com itraconazolou cetoconazol é eficaz.

3.1.2.5. Papilomatose Confluente e Reticulada

A papilomatose confluente e reticulada (Gougerot, Carteaud, 1927) é uma condição rara que afeta o tronco superior e os pacientes desenvolvem uma rede de lesões lineares descamativas não inflamatórias (Abdillah, Ranque, 2021). Biópsias das lesões podem revelar a presença de leveduras *Malassezia*, mas as evidências de que elas desempenham um papel na causa dessa condição são insuficientes. Em primeiro lugar, não há uma explicação patogenética satisfatória nem evidências sólidas de que as lesões respondam aos antifúngicos. A condição é muito incomum e, portanto, estudos terapêuticos comparativos controlados são difíceis de serem realizados. No entanto, em geral, os pacientes respondem à minociclina, o que também é difícil de explicar com base em um papel central para as espécies de *Malassezia*.

3.1.2.6. Outras Condições

Espécies do gênero *Malassezia* tem sido implicadas em outras condições, como blefarite e otite externa (Xu, Li, 2019; Rhimi et al., 2020; Saunte et al., 2020; Ianiri, Heitman, 2020; Abdillah, Ranque, 2021; Park et al., 2021; Hobi et al., 2022; Kurniadi et al., 2022). Novamente, o problema reside em relacionar a patogênese aos organismos presentes no nicho anatômico, uma vez que as leveduras lipofílicas podem ser encontradas tanto nos cílios quantono conduto auditivo externo. Além disso, a blefarite e a descamação da orelha externa são observadas em pacientes com dermatite seborreica, e a ligação pode ser semelhante àquela vista em pacientes com dermatite seborreica. Existe uma forma de psoríase, às vezes chamada de sebopsoríase, na qual a distribuição clínica das lesões pode se assemelhar às observadas na dermatite seborreica. Pacientes com essa forma de psoríase relataram melhora com antifúngicos, sugerindo assim que a *Malassezia* (ou outro fungo) possa desempenhar um papel nessa doença (Rosenberg, Belew, 1982). No entanto, não está claro se isso é uma entidade distinta ou se representa uma sobreposição entre dermatite seborreica e psoríase - ou até mesmopsoríase co-localizada (efeito de Koebner) com dermatite seborreica (Narang et al., 2007).

Raramente, as espécies de *Malassezia* têm sido descritas como causas de infecção sistêmica, geralmente em neonatos. O primeiro caso, em 1981, relatou uma associação com o uso de alimentação intravenosa lipídica e cuidados em uma UTI neonatal (Redline, Dahms, 1981). A transmissão por parte da equipe médica, outros pacientes ou seus parentes foi postulada (Sizun et al., 1994). A causa usual é *M. pachydermatis*. Agrupamento de casos foi relatado, sugerindo o potencial de transmissão múltipla. Essa condição é fatal se não tratada, e os pacientes respondem, se tratados em tempo hábil com antifúngicos sistêmicos. No entanto, dado que se trata de um patógeno incomum, o problema muitas vezes é o diagnóstico tardio dessa infecção.

3.1.3 *Malassezia*: taxonomia

A descrição mais antiga para o gênero *Malassezia* remonta a 1846, quando Eichstedt reconheceu a natureza fúngica da doença, pitiríase versicolor (PV) (Eichstedt, 1846). Ele observou a presença de leveduras e filamentos em material de pacientes, mas o organismo só recebeu um nome quando Robin (1953) o identificou como *Microsporon furfur*. Esse nome foi substituídopor Baillon (1889), que criou um novo gênero, Malassezia, estabelecendo assim o nome binomial *Malassezia furfur*. Relatos contemporâneos dessa época revelaram um interesse em condições escamosas do couro cabeludo, incluindo dermatite seborreica (DS) e caspa (pitiríase capitis), nas

quais a presença de leveduras sem filamentos foi observada não apenas nesses pacientes, mas também em pele aparentemente saudável. Rivolta (1873) observou organismos de contorno duplo em escamas de lesões de psoríase, classificando-os dentro do gênero *Cryptococcus*, e no ano seguinte, Malassez (1874) descreveu "esporos" esféricos e ovais em lesões escamosas do couro cabeludo. Esses esporos foram nomeados como duas espécies de *Saccharomyces*, *S. ovalis* e *S. sphaericus*, por Bizzozero (1884), mas outros pesquisadores consideraram que eles representavam uma única espécie (Unna, 1891; Sabouraud, 1904), e foi Sabouraud (1904) quem os atribuiu a um novo gênero, *Pityrosporum*. Embora ele reconhecesse as semelhanças entre PV e pitiríase capitis, tanto nos sinais clínicos quanto na localização superficial de colonização pelas leveduras, a presença de filamentos nas escamas de PV e sua ausência em material de outras condições levou à relutância em colocar *M. furfur* e *Pityrosporum* em um único gênero (Slooff, 1970). Portanto, as leveduras observadas em couros cabeludos escamosos foram nomeadas por ele como *Pityrosporum malassezii*, mas isso foi posteriormente alterado para *P.ovale* por Castellani e Chalmers (1913), que reconheceram o nome específico anteriormente atribuído por Bizzozero (1884).

Estudos subsequentes reconheceram semelhanças entre *Malassezia* e *Pityrosporum*, sugerindo um único gênero para essas leveduras, nomeando-as *M. furfur* e *M. ovalis*, com umaterceira espécie, *M. tropica*, como causa de uma variante tropical de PV, a pitiríase flava (Panja, 1927). No entanto, essa ideia não foi amplamente aceita na época, e continuou sendo conveniente utilizar o nome *M. furfur* para descrever a característica microscópica de diagnóstico das leveduras juntamente com os filamentos em escamas de PV, e continuar usando*Pityrosporum* para identificar as leveduras encontradas em outras condições e em estudos de cultura. Embora várias combinações dessas leveduras tenham aparecido na literatura até a década de 1930, a maioria dos autores manteve o nome binomial *Pityrosporumovale* (Ingham, Cunningham, 1993). Essa espécie, com seus vários nomes, continuou a incluir leveduras com morfologia diversa, incluindo células esféricas (Sabouraud, 1904; Unna, 1891; Ota, Huang, 1933).

Durante os anos de 1925 a 1955, foram descritas mais espécies de *Pityrosporum*. Em 1925, Weidman isolou leveduras da pele de um rinoceronte, que se assemelhavam a *P. ovale* em morfologia, mas podiam ser distinguidas pela falta de dependência de lipídios para crescimento (Weidman, 1925). Ele a chamou de *P. pachydermatis*. Uma proposta de Dodge (1935) para adotar a designação *Malassezia pachydermatis* para essa levedura não foi aceita na época. Uma levedura semelhante foi descrita por Gustafson (1955) a partir de otite externa em cães, que ele chamou de *P. canis*, mas isso posteriormente se

tornou sinônimo de *P. pachydermatis* (Slooff, 1970; Fraser, 1961). Essa espécie não é um membro normal da micobiota humana, mas foi encontrada em uma variedade de hospedeiros animais.

Com o avanço das características moleculares, ficou evidente que as variedades relatadas na literatura representavam uma diversidade deespécies dentro do gênero *Malassezia*. Uma nova espécie, *M. sympodialis*, foi descrita em 1990, sendo diferenciada por experimentos de reassociação de DNA/DNA e valores de conteúdo de G+C do DNA nuclear (Simmons, 1990). Essa espécie foi incorporada ao gênero na quarta edição de "*The Yeasts*" (Boekhout et al., 1993). Posteriormente, foram estabelecidas quatro novas espécies em 1996 e, assim, os dados moleculares puderam ser correlacionados com propriedades fisiológicas, como a capacidade de utilizar diferentes suplementos lipídicos, diferenças na reação da catalase e variação na tolerância à temperatura (Guého et al., 1996).

Foram desenvolvidos esquemas que permitiram a identificação das espécies com base em características fenotípicas (Crespo Erchiga e Guillot, 1996; Guillot et al., 1996), o que levou a publicações sobre estudos ecológicos mostrando a distribuição das espécies de acordo com condições patológicas e áreas geográficas (Crespo Erchiga et al., 1999; Deesi, 2006; Gupta et al., 2001).

Atualmente, as metodologias de biologia molecular têm sido consideradasde grande importância para o agrupamento taxonômico do gênero *Malassezia*. O apoio de ferramentas moleculares é muito importante, pois as características morfológicas entre as espécies são muitosutis, o que pode gerar dúvidas e erros. O estudo das regiões D1/D2 do gene do RNA ribossômico 28S (rRNA28S) e das regiões ITS (*Internal Transcribe Spacer*) têm auxiliado na diferenciação das espécies de *Malassezia* (Batra *et al.*, 2005). Em 1996, com base em estudos baseados em morfologia, fisiologia e biologia molecular, houve uma mudança na classificação do gênero *Malassezia* em sete espécies. A partir de 2002, outras sete novas espécies de *Malassezia* foram citadas na literatura. Como descrito anteriormente, 18 espécies são reconhecidas atualmente para o gênero (Vijaya Chandra *et al.*, 2021), sendo que apenas 17 possuem genomas disponíveis (*Malassezia sp., M. arunalokei, M. caprae, M. cuniculi, M. dermatis, M. equina, M. furfur, M. globosa, M. japonica, M. nana, M. obtusa, M. pachydermatis, M. restrita, M. slooffiae, M. sympodialis, M. vespertilionis* e *M. yamatoensis*) e 9 com genomas mitocondriais (*M. furfur, M. globosa, M. japonica, M. obtusa, M. pachydermatis, M. restrita, M. slooffiae, M. sympodialis* e *M. yamatoensis*).

4. ARTIGO I

Este capítulo contempla os seguintes objetivos: **(1)** identificar os genomas mitocondriais do gênero *Malassezia* já disponíveis nos bancos de dados do NCBI; **(2)** identificar as bibliotecas de sequenciamento de genoma completo (WGS) disponíveis para o gênero *Malassezia* na base de dados SRA do NCBI; **(3)** implementar um fluxo de trabalho de bioinformática para a montagem e anotação de novos genomas mitocondriais de fungos do gênero *Malassezia*; **(4)** anotação funcional dos novos mitogenomas; **(5)** análise filogenética dosnovos mitogenomas em conjunto com os já disponíveis.

Artigo publicado na RevistaFT (Qualis B2):
Disponível: Huszcz G.B; Teixeira G.E; Rodrigues-Oliveira I.H; Rocha R.R; Oliveira RC;Takao PS; Aciole Barbosa D.; Jabes DL; Kavalco KF; Pasa R; Menegidio F.B. Five mitochondrial genomes of the lipophilic fungus *Malassezia* (*Malassezia*les, Exobasidiomycetes). RevistaFT, v. 27, n. 124, p. 1-2, 2023. https://doi.org/10.5281/zenodo.8140802

Título: Five mitochondrial genomes of the lipophilic fungus *Malassezia* (*Malassezia*les, Exobasidiomycetes)

Gabriel Barbosa Huszcz[1,2], Gledison Eric Teixeira[1], Igor Henrique Rodrigues-Oliveira[3], RenanRodrigues Rocha[2,3], Rafaela de Campos Oliveira[2], Paulo Sallarola Takao[2], David Aciole Barbosa[1], Daniela Leite Jabes[1], Karine Frehner Kavalco[3], Rubens Pasa[3], Fabiano Bezerra Menegidio[1,2]

[1]Integrated Biotechnology Center, University of Mogi das Cruzes, Mogi das Cruzes, SP,Brazil.

[2] Technological Research Center, University of Mogi das Cruzes, Mogi das Cruzes, SP, Brazil.

[3] Laboratory of Ecological and Evolutionary Genetics, Federal University of Viçosa, Rio Paranaíba, MG, Brazil.

* Corresponding Author: fabianomenegidio@umc.br

Five mitochondrial genomes of the lipophilic fungus *Malassezia* (*Malasseziales*, *Exobasidiomycetes*)

Abstract: Fungi of the *Malassezia* genus are the most prevalent eukaryotic organisms in the natural skin mycobiome of animals and humans. The genus is found in various parts of the body, causing common skin diseases such as pityriasis versicolor and seborrheic dermatitis, and its high adaptability to the human body and prevalence among other eukaryotes makes it unique. However, even with eighteen known species and genomes available, studies on their phylogenyand basic genomic information are still scarce. To this day, only nine species have their mitochondrial genome described and assembled. In this study, we describe for the mtDNA of *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina*, *Malassezia caprae* and six new mitogenomes of *Malassezia pachydermatis*. In addition, to better understand the phylogenetic relationships, we mined 35 mitogenomes belonging to 9 species of *Malassezia*, along with the mitogenomes gathered in this study, and reconstructed the phylogeny based on mtDNA. The results of this study provide genomic variation information and enhance the understanding of the *Malassezia* genus. They could later deliver highly valuable new insight into data for phylogenetic analysis and population genetics.

Keywords: Mitochondrial genome, *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina*, *Malassezia caprae*, *Malassezia pachydermatis*,phylogeny.

Introduction

Yeasts of the *Malassezia* genus are present in the microbiota of many animals, including humans, living as commensal organisms on the skin and scalp (Findley *et al.*, 2020). However, in some cases, they can become pathogenic under the influence of predisposing factors that involve both the conditions of the skin microenvironment and changes in the host's immune system (Theelen *et al.*, 2018; Aykut *et al.*, 2019; Limon *et al.*, 2019; Saunte *et al.*, 2020; Spatz and Richard 2020). The genus groups basidiomycetic, lipophilic (*M. pachydermatis*) and/or lipodependent yeasts that show asexual reproduction by unipolar or sympodial budding (*M. sympodialis*) from the parent cell. Another important feature is the possibility of forming pseudohyphae.

Differentiation of species in the traditional way is done through lipid and fatty acid assimilation tests (biochemical methods), also considering the characteristics of micromorphology. Currently, molecular biology methodologies have been considered of great importance for the taxonomic grouping of the *Malassezia* genus. The support of molecular toolsis very important, since the morphological characteristics between species are very subtle, which can generate doubts and errors. Study of the D1/D2 regions of the 28S ribosomal RNA gene (rRNA28S) and the ITS (Internal Transcribed Spacer) regions have helped in the differentiation of *Malassezia* species (Batra *et al.*, 2005). In 1996, based on studies based on morphology, physiology and molecular biology, there was a change in the classification of thegenus *Malassezia* in seven species. As of 2002, another seven new species of *Malassezia* were cited in the literature. Currently, 18 species are recognized for the genus, with only 17 having available genomes (*M. arunalokei*, *M. caprae*, *M. cuniculi*, *M. dermatis*, *M. equina*, *M. furfur*, *M. globosa*, *M. japonica*, *M. nana*, *M. obtusa*, *M. pachydermatis*, *M. restricta*, *M. slooffiae*, *Malassezia sp.*, *M. sympodialis*, *M. vespertilionis* and *M. yamatoensis*) and 9 having mitochondrial genomes (*M. furfur*, *M. globosa*, *M. japonica*, *M. obtusa*, *M. pachydermatis*, *M. restricta*, *M. slooffiae*, *M. sympodialis* and *M. yamatoensis*).

In this study, we describe the mitochondrial genome of five species of *Malassezia*les: *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina* and *Malassezia caprae*, also six new mitogenomes of *Malassezia pachydermatis*. In addition, to better understand the phylogenetic relationships, we mined 35 mitogenomes belonging to 9 species of *Malassezia*, along with the mitogenomes gathered in this study, and reconstructed the phylogeny based on mtDNA.

Materials & Methods

Mitogenome assembly and annotation

We obtained Whole Genome Sequencing (WGS) raw dataof *Malassezia vespertilionis* (SRR6206152; Lorch *et al.*, 2018), *Malassezia dermatis* (DRR043255; Sugita *et al.*, 2002), *Malassezia nana* (DRR043258; Hirai *et al.*, 2004), *Malassezia equina* (SRR2136627; Wu *et al.*, 2015) and *Malassezia caprae* (SRR2132347; Wu *et al.*, 2015) from the Sequence Reads Archive (SRA) NCBI.

The reads were submitted to a workflow previously used and validated in fish mitogenome assemblies by Rocha-Reis *et al.* (2020), Resende *et al.* (2020) and Pasa *et al.* (2021), with minor modifications to adapt to the assembly of fungal mitogenomes. We imported the raw data into the Galaxy Europe platform (https://usegalaxy.eu/) and usedNOVOplasty v4.3.1 (Dierckxsens *et al.*, 2017) to assemble the mitochondrial genomes. As a seed, we used the complete mitogenome of *M. japonica* (KY911090.1). We annotated the sequences obtainedon Mitos2 (http://mitos2.bioinf.uni-leipzig.de/; Arab *et al.*, 2017) using Gene Code 3 - Yeast and RefSeq 89 Fungi.

To validate our workflow, we also performed the assembly and annotation of six new mitogenomes from *Malassezia pachydermatis* WGS libraries (SRR12046913, SRR12005311, SRR12005312, SRR12005313, SRR12005314 and SRR2135029). In addition, the 34 mitogenomes of the 9 species of the *Malassezia* genus available at the NCBI (*M. furfur*, *M.globosa*, *M. japonica*, *M. obtusa*, *M. pachydermatis*, *M. restricta*, *M. slooffiae*, *M. sympodialis* and *M. yamatoensis*) were annotated according to the procedure described above.The circular map of the new mitogenomes was drawn using the online software GenomeVx (Conant and Wolfe, 2008). We used the Blast Ring Image Generator (BRIG)(Alikhan *et al.*, 2011) to perform a comparative Blast analysis of available representative *Malassezia* mitogenomes against our mitogenome pools.

Mitochondrial phylogenetic analysis

To perform the phylogenetic analyses, we manually extracted the sequences of the 15protein-coding genes (PCGs) of the 11 mitogenomes assembled in this work and added to the dataset the PCGs of the 34 mitogenomes of other *Malassezia* species available at the NCBI, addition to the *Rhodotorula mucilaginous* (MF694646.1) mitogenome as an outgroup. We aligned the sequences with MAFFT v7 (Katoh & Standley, 2013), available on the Galaxy Europe online platform (Afgan *et al.*, 2018), and concatenated the alignments in Sequence Matrix v1.8 (Vaidya *et al.*, 2011). We constructed the phylogeny by the Maximum Likelihood (ML) method in the IqTree v2.1.2 software (Nguyen *et al.*, 2015) using 1000 ultra-fast bootstrap repeats and the evolutionary model GTR+F+I+G4 estimated by the IqTree Model Finder (Kalyaanamoorthy *et al.*, 2017).

Data Description

Characterization of mitochondrial genomes of *Malassezia* species

The mitogenome of the five *Malassezia* species are circular DNA molecules with a total length ranging from 27,284 bp to 40,970 bp. Mitogenome sizes varied widely among the five newly assembled mitochondrial genomes, with *M. vespertilionis* having the largest among the new mtDNA and *M. caprae* being the smallest mitogenome ever described for the genus. Furthermore, *M. equina* becomes the second smallest mitogenome described for the genus, followed by the already described *M. japonica*. However, *M. dermatis* and *M. nana* presented a mitogenome with a size close to the average of other species of the genus. Likewise, the new
M. pachydermatis mitogenomes were between 35,817 bp and 35,829 bp in size, being slightly larger than the *M. pachydermatis* strain CBS1879 (KY911092.1) mitogenome previously described. A comparison between the size of the new mitogenomes described in this work and those available in the NCBI database can be seen in Supplementary Tables 1.

The GC concentration of the mitogenome differs among species, which may be affected by mutation bias, selection, and biases of reconstitution-related DNA repair (Li *et al.*, 2018). The low G + C content observed is similar to other fungal mitogenomes in *Malassezia* (Supplementary Table 1). The GC content in the five mitogenomes ranged from 23.4% to 33.7%, among which the GC content in the genome of *M. caprae* was the highest.

The base composition of the mitochondrial genome of *M. caprae* was estimated as 33.8% A, 17.2% C, 16.5% G, and 32.5% T, thus presenting a bias towards AT (66.3%) and theGC content was 33.7%. *M. equina* presented the composition of 33.4% A, 33.7% T, 16.0% C and 16.1% G, with a bias towards AT (67.1%) and GC content of 32.9%. *M. nana* presented the composition of 33.6% A, 34.5% T, 15.8% C and 16.1% G, with a bias towards AT (68.1%) and GC content of 31.9%. *M. vespertilionis* presented the composition of 36.8% A, 39.8% T, 11.5% C and 11.8% G, with a bias towards AT (76.6%) and GC content of 23.4%. Finally, *M. dermatis* presented a composition of 34.1% A, 34.3% T, 15.6% C and 16.0% G, with a bias towards AT (68.4%) and GC content of 31.6%.

In general, mitogenomic coding is associated with the mitochondrial translation apparatus, electron transport, and oxidative phosphorylation. The potential coding

genes(PCGs) of *Malassezia* species included 14 PCGs related to oxidative phosphorylation [NAD dehydrogenase (nad1-6 and nad4L), cytochrome oxidases (cox1, cox2 and cox3), cytochrome b (cob), and 3 ATP synthases (atp6, atp8 and atp9), ribosomal protein S3 (rps3)] and 2 ribosomal RNA genes (rns and rnl). Only *M. equina*, *M. caprae* and *M. vespertilionis* presented 14 PCGs, due to the absence of the atp9 gene. This distribution is consistent with that observedin typical fungal mitogenomes, such as those belonging to *Ascomycota* and *Basidiomycota* (*Dikarya* Subkingdom). In total, the coding regions containing the 15 PCGs described occupiedbetween 13,725 bp and 14,103 bp in the two mitogenomes presented (*M. nana* and *M. dermatis*,respectively). Likewise, the 15 PCGs occupied from 14,037 bp to 14,208 bp among the 6 new mitogenomes of *M. pachydermatis* described here, in addition to 12,735 bp and 14,784 bp among the mitogenomes available in the NCBI database. In mitogenomes that presented 14 PCGs, the coding regions occupied between 13,437 bp and 14,079 bp (*M. equina* and *M. caprae*, respectively). Other existing regions in mitogenomes are tRNAs, rRNA genes and intergenic regions. More details on these regions contained in mitogenomes are available in Supplementary Table 1.

Except for *M. vespertilionis*, all new mitogenomes showed nad4 and atp6 in the light chain. Likewise, a duplication of the atp9 gene in *M. dermatis*, *M. nana* and in the new *M. pachydermatis* mitogenomes was also observed in the light chain. In *M. vespertilionis*, only the cob gene was observed in the light chain. Apart from those already described, all other genes were observed in the heavy chain for all new mitogenomes. More details on thedirection of PCGs can be found in Supplementary Tables 1-2.

Besides *M. caprae* and *M. equina*, where 18 tRNA genes were identified in the mitogenomes, the other species characterized in this work presented between 20 and 25 tRNAs,encoding the 20 standard amino acids. The length of tRNA genes in the five mitogenomes ranged from 71 bp to 86 bp. Each of the five *Malassezia* mitogenomes involved 2 rRNA genes:the small subunit ribosomal RNA gene (rns) and the large subunit ribosomal RNA gene (rnl). No significant difference was identified in the number of rRNA genes in the five species, and the total length of the genes.

Phylogenetic Analysis

We obtained a robust evolutionary tree topology (Figure 1) with all recovered clades well supported. Our phylogeny is mostly accordant with a prior study based on sequences of six mitochondrial and nuclear genes (Wang *et al.*, 2014). The tree has shown two main clades comprising a distinct group of species. As in the literature, our data

reached a clade encompassing *M. caprae, M. dermatis, M. sympodialis, M. equina,* and *M. nana*. This clade is near related to *M. pachydermatis* clade. Another well-marked and related clade comprises *M. globosa* and *M. restricta* lineages.

The other main clade assembles *M. sloffiae, M. vespertilionis, M. japonica, M. yamatoensis*, *M. obtusa*, and a species complex formed by samples identified as *M. furfur*. Ouranalysis shows two clades of *M. furfur* with 100% bootstrap confidence and a single strain diverging early. The bootstrap confidence of this strain with the two largest clades reached 100% of confidence.

The main difference between our analysis and Wang *et al.* (2014) study was the position of *M. restricta*, *M. globosa*, and *M. pachydermatis*. In our three (Figure 1), these species are closest to the first mentioned clade, though previous analyzes grouped them with the second clade (Wang *et al.*, 2014). In the phylogenetic analysis, the novel *Malassezia pachydermatis* mitogenome had clustered consistently with *M. pachydermatis* strainCBS1879(KY911092.1). The same had happened with other species with multiple mitogenomes available.

Conclusion

The methodology used to assembly the mitochondrial genomes of *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina* and *Malassezia caprae* proved to be satisfactory and allowed the expanded the phylogeny of *Malasseziales* through mtDNA. We were also able to validate our workflow by assembling 6 new *Malassezia pachydermatis* mitogenomes that demonstrated high agreement with the mitogenome already available. Furthermore, the study of the complete mitochondrial genome proves to be a tool with potential to solve taxonomic problems and help in understanding the evolutionary and ecological relationships of this group.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercialor financial relationships that could be construed as a potential conflict of interest.

Author Contributions

RCO, IHRO, GH, GET, PST, RRR, DAB, DLJ, KFK, RP and FBM contributed to the conceptualization, investigation and design of the study. RCO, IHRO, GH, GET, PST, RRR,DAB, DLJ, KFK, RP and FBM performed the bioinformatics/statistics analysis and were responsible for data validation. FBM and DLJ provided resources to carry out the work. FBM,DLJ, KFK and RP were responsible for the administration and supervision of the research project. FBM, DAB, RCO, IHRO, PST and RRR wrote the first draft of the manuscript. FBM, DLJ, KFK and RP wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Acknowledgments

This study was financed in part by scholarship grants from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil - CAPES (www.capes.gov.br/) and Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (www.gov.br/cnpq/pt-br).

Supplementary Material

The Supplementary Material for this article can be found online at:https://doi.org/10.6084/m9.figshare.23666289.v1

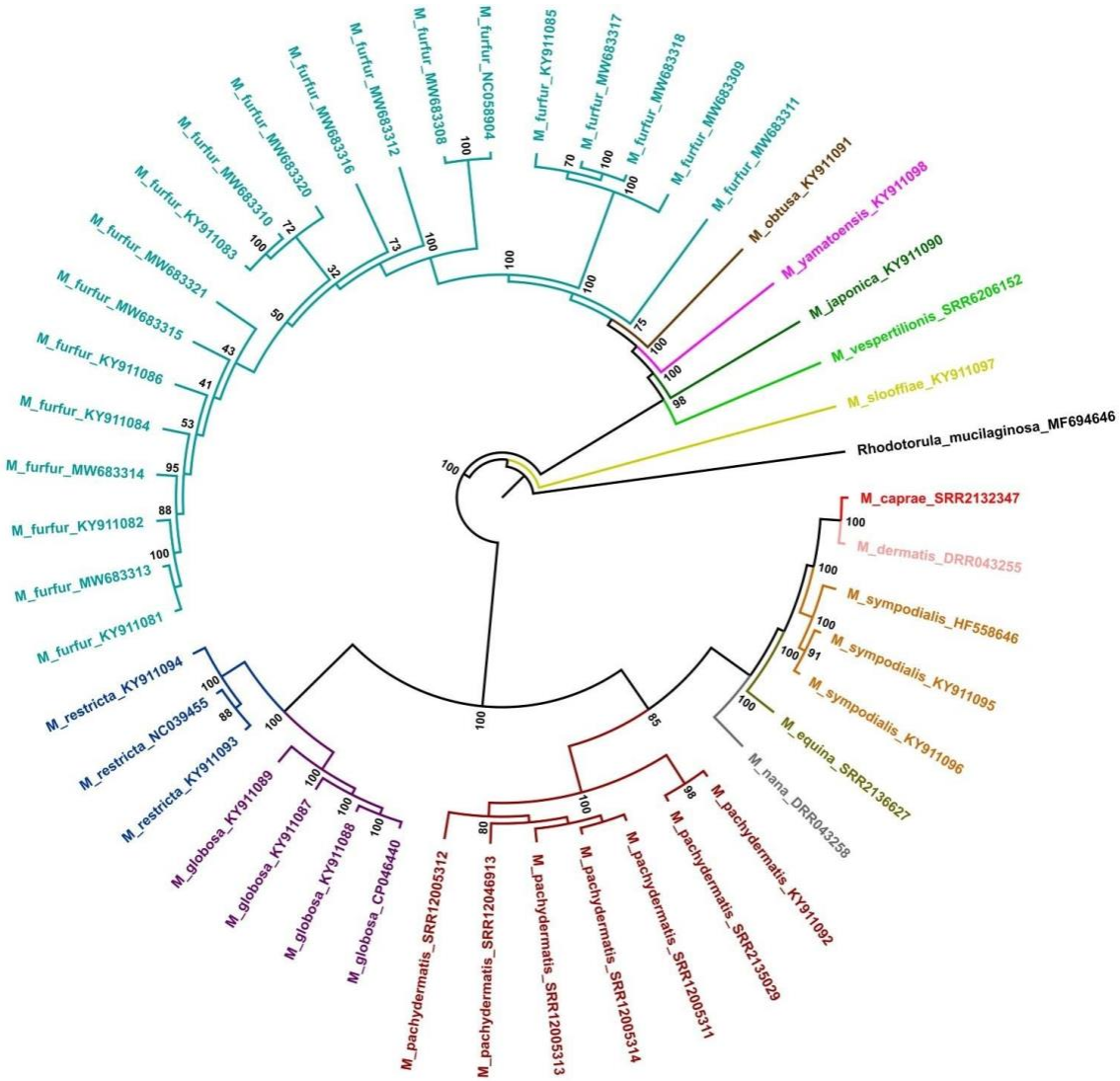
Reference

1. Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M *et al.* The Galaxyplatform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic acids research. 2018;46(W1):W537-W544.
2. Al Arab M, Zu Siederdissen CH, Tout K, Sahyoun AH, Stadler PF, Bernt M. Accurateannotation of protein-coding genes in mitochondrial genomes. Molecular phylogenetics and evolution. 2017;106: 209-216.
3. Alikhan, NF., Petty, N.K., Ben Zakour, N.L. *et al.* BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12, 402 (2011).https://doi.org/10.1186/1471-2164-12-402
4. Aykut B, Pushalkar S, Chen R *et al.* The fungal mycobiome promotes pancreaticoncogenesis via activation of MBL. Nature. 2019;574:264–7.
5. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic acids research. 2017;45(4):e18-e18.
6. Findley K, Oh J, Yang J *et al.* Topographic diversity of fungal and bacterialcommunities in human skin. Nature. 2020;498:367–70.
7. Hirai A, Kano R, Makimura K, Duarte ER, Hamdan JS, Lachance MA, Yamaguchi H,
   Hasegawa A. *Malassezia* nana sp. nov., a novel lipid-dependent yeast species isolatedfrom animals. Int J Syst Evol Microbiol. 2004 Mar;54(Pt 2):623-627. doi: 10.1099/ijs.0.02776-0. PMID: 15023986.
8. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder:
   fast model selection for accurate phylogenetic estimates. Nature Methods2017;4(6):587-589.
9. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013;30(4):772–780.
10. Limon JJ, Tang J, Li D *et al. Malassezia* is associated with Crohn's disease andexacerbates colitis in mouse models. Cell Host Microbe. 2019;25:377–88.
11. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference inthe genomic era. Molecular Biology and Evolution 2020;37(5):1530-1534.
12. Saunte DML, Gaitanis G, Hay RJ. *Malassezia*-Associated skin diseases, the use ofdiagnostics and treatment. Front Cell Infect Microbiol. 2020;10:112.
13. Spatz M, Richard ML. Overview of the potential role of *Malassezia* in gut health anddisease. Front Cell Infect Microbiol. 2020;10:201.
14. Theelen B, Cafarchia C, Gaitanis G *et al. Malassezia* ecology, pathophysiology, andtreatment [published correction appears in Med Mycol. 2019 Apr 1;57(3):e2]. Med Mycol. 2018;56:S10–25.
15. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fastassembly of multi-gene datasets with character set and codon information. Cladistics 2011;27(2):171-180.
16. Wang QM, Theelen B, Groenewald M, Bai FY, Boekhout T. Moniliellomycetes and Malasseziomycetes, two new classes in Ustilaginomycotina. Persoonia-MolecularPhylogeny and Evolution of Fungi. 2014;33(1):41-47.

17. Wu, G., Zhao, H., Li, C., Rajapakse, M. P., Wong, W. C. *et al.* Genus-wide comparative genomics of *Malassezia* delineates its phylogeny, physiology, and nicheadaptation on human skin. PLoS Genetics 11, e1005614 (2015)

FIGURE

Figure 1. Phylogeny of the mitogenomes of the genus *Malassezia*



5. ARTIGO II

Este capítulo contempla os seguintes objetivos: (1) realizar a reanotação funcional de genomas mitocondriais em uma variedade de espécies de fungos do gênero *Malassezia*; (2) organizar as anotações obtidas em um conjunto de dado reutilizável e de fácil entendimento;
(3) construir e implementar um banco de dados relacional baseado no projeto *Machado*, utilizando o esquema *Chado* dentro do *PostgreSQL SGDB*; (4) popular as tabelas do banco de dados com dados genômicos e funcionais dos mitogenomas de *Malassezia* obtidos nas etapas anteriores; (5) implementar uma aplicação web baseada no *SequenceServer* que permita pesquisas BLAST nas sequências genômicas de *Malassezia*.

Artigo submetido para a Revista Latin American Data in Science.
Disponível: https://ojs.datainscience.com.br/

Título: Mining mitochondrial genomes of *Malassezia* fungi: building a comprehensive dataset and database

Gledison Eric Teixeira[1], Gabriel Barbosa Huszcz[1,2], David Aciole Barbosa[1], Igor Henrique Rodrigues-Oliveira[3], Renan Rodrigues Rocha[2,3], Karine Frehner Kavalco[3], Rubens Pasa[3], Fabiano Bezerra Menegidio[1,2]

[1]Integrated Biotechnology Center, University of Mogi das Cruzes, Mogi das Cruzes, SP,Brazil.
[2] Technological Research Center, University of Mogi das Cruzes, Mogi das Cruzes, SP,Brazil.
[3] Laboratory of Ecological and Evolutionary Genetics, Federal University of Viçosa, Rio Paranaíba, MG, Brazil.

* Corresponding Author: fabianomenegidio@umc.br

Mining mitochondrial genomes of *Malassezia* fungi: building a comprehensive datasetand database

Abstract: The study of mitochondrial genomes in *Malassezia* fungi holds significant importance in understanding their genetic diversity and functional characteristics. In this research, we present a comprehensive approach to mining and analyzing mitochondrial genomes of various *Malassezia* species. Our primary objective was to build a comprehensive dataset that encompasses a wide range of genomic information from these fungi. Leveraging this dataset, we implemented a relational database that serves as a powerful resource for storing, managing, and querying mitochondrial genomic data. Through a combination of bioinformatictools, we conducted functional annotation of the mitochondrial genomes, identifying crucial genes, regulatory elements, and potential functional pathways. This annotation process was incorporated into the database, enhancing its utility for researchers and facilitating cross- species comparisons. The constructed dataset and database provide an invaluable resource for researchers studying *Malassezia* fungi and their mitochondrial genomes. The database's user- friendly interface allows for easy access to genomic data, enabling efficient exploration and analysis. Additionally, the dataset opens avenues for in- depth investigations into the evolutionand adaptation of these fungi to diverse ecological niches. This article highlights the significance of a comprehensive dataset and database in the field of mitochondrial genomics and its potential to advance our knowledge of *Malassezia* fungi. The integration of functional annotations with the genomic data will foster new discoveries and contribute to broader studieson fungal biology and evolution.

Keywords: *Malassezia*; mitogenomes; relational database; Machado database.

Data importance

- *Malassezia* fungi comprise a diverse group of species that inhabit diverse environments,including the skin of humans and other animals (Findley *et al.*, 2020). Some members

  of this genus have been linked to skin disorders, while others play crucial roles in industrial processes, such as lipid biotransformation (Theelen *et al.*, 2018; Aykut *et al.*,2019; Limon *et al.*, 2019; Saunte *et al.*, 2020; Spatz and Richard 2020).

However, despite their significance, a comprehensive understanding of the mitochondrial genomes of *Malassezia* fungi is still lacking. The primary objective of this study is threefold: to perform functional annotation of mitochondrial genomes in a variety of *Malassezia* fungal species, unveiling their genetic features and functional elements. This comprehensive dataset, comprising twenty *Malassezia* mitogenomes, was compiled through state-of-the-art high- throughput sequencing and extensive bioinformatic analysis. The resulting insights into vital mitochondrial genes, regulatory regions, and functional pathways shed light on their adaptation and biological roles in these fungi.

- Secondly, we have successfully constructed a relational database based on the project Machado (Mudadu and Zerlotini, 2020), utilizing the Chado schema within the PostgreSQL SGDB (Mungall and Emmert, 2007). This database server as a centralized repository for researchers interested in exploring and analyzing mitochondrial genomes of *Malassezia* fungi, facilitating cross-species comparisons, and enabling future investigations. With the database readily available, researchers canefficiently access and utilize the data to delve deeper into the genetic intricacies of these fungi, accelerating the pace of scientific discoveries and advancements in various fields of fungal genomics. The adoption of the Chado schema ensures a robust and standardized structure, enhancing the database's utility and interoperability within the scientific community.
- Lastly, as a significant achievement, we have successfully developed a web application based on the Sequence Server framework (Priyam *et al.*, 2019), integrated with the service of containers Docker (https://www.docker.com/). This user-friendly web application provides an intuitive interface for researchers to access and explore the comprehensive genomic data stored in our database. Leveraging the power of Docker containers ensures easy deployment and scalability, allowing seamless access to the application across various platforms and environments. The implementation of this web application represents a major step forward in enhancing the accessibility and usability of the genomic data, empowering researchers to delve deeper into the geneticin tricacies of these fungi and further accelerating scientific discoveries in the field of fungal genomics.

Materials and Methods

Data Collection and Preprocessing

For this work, we obtained a total of 40 mitochondrial genomes belonging to 10 species of fungi from the genus *Malassezia*. These genomes were retrieved from the Nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore) and/or the Organelle Genome Resources (https://www.ncbi.nlm.nih.gov/genome/organelle/) maintained by the National Center for Biotechnology Information (NCBI). Additionally, during the compilation of the dataset and database, we incorporated 5 mitochondrial genomes presented in the work of Huszcz *et al.* (2023).

To ensure the accuracy and integrity of the downloaded files, a thorough data integritycheck was carried out. This step aimed to detect any corruption or issues that might have occurred during the download process.

All relevant metadata associated with the mitochondrial genomes, including species information, DNA sequences, annotations, and other details, were then extracted from the downloaded files and meticulously documented for future reference.

To maintain consistency in the structure of the data, the obtained mitochondrial genomes were standardized. This involved ensuring uniformity in file formats and sequence nomenclature across all genomes.

Functional Annotation

The functional annotation of the mitochondrial genomes was performed using Mitos2(http://mitos2.bioinf.uni-leipzig.de/; Arab *et al.*, 2017) in conjunction with Gene Code3 - Yeastand RefSeq 89 Fungi databases. This annotation process aimed to identify genes encoding essential mitochondrial proteins, regulatory regions, and potential functional pathways within the genomes.

Implementation of the Machado Framework

In this study, we successfully implemented the Machado Framework (Mudadu and Zerlotini, 2020) on a server to establish a robust infrastructure for the database. The implementation process comprised several crucial steps, adhering to instructions available in the research article and the project's GitHub repository.

To begin, we carefully followed the official Docker instructions to install Docker infrastructure (https://docs.docker.com/engine/install/ubuntu/) and Docker Compose tool

(https://docs.docker.com/compose/) on the server. These tools are essential for the smooth operation of the Machado Framework and the efficient execution of the required containers.

Furthermore, to optimize container construction and minimize image size, we performed a refactoring of the provided Dockerfile. By reducing the use of the RUN parameter, we minimized the number of layers generated during container construction.Subsequently, we created a new Docker Compose file, which referred to our updated Machado Framework image.

With the server ready and the Docker containers properly configured, we proceeded toimplement the Machado Framework environment by cloning the Machado Framework repository. As part of this setup, we created directories for PostgreSQL (https://www.postgresql.org/), ElasticSearch (https://www.elastic.co/what-is/elasticsearch), and JBrowse (Diesh *et al.*, 2023) data, and carefully edited the Machado environment variablesin the .env file.

Upon replacing the original Dockerfile and docker-compose.yml files with our refactored versions, we executed the Machado environment using Docker Compose. This deployment enabled convenient access to the web interface of the Machado Framework at http://hostname:8000/machadosample/.

Overall, this comprehensive implementation allowed us to offer a powerful and user- friendly platform, empowering researchers to explore and analyze mitochondrial genomes of various *Malassezia* fungal species. This significant advancement fosters progress in the field of fungal genomics and holds great potential for broader applications in various scientific domains.

Implementation of the Sequence Server

To facilitate access to the mitochondrial genomes, we provided a database based on the SequenceServer (Priyam *et al.*, 2019) that allowed the search for sequence similarity through the BLAST tool (Altschul *et al.*, 1990). This tool displays alignment results in visual graphics which can be downloaded in SVG and PNG formats. It also provides direct downloadof sequence and alignment from the result page.

Data Description

Malassezia mitogenome dataset

Resulting from this research, a comprehensive General Feature File (GFF3) of ~13 kilobytes has been generated and is available for retrieval from the Open Science Framework (OSF) repository (see supplementary materials). This GFF3 file contains valuable data concerning protein coding genes (PCGs), transfer RNAs (tRNAs), ribosomal RNA (rRNAs), mtDNA control region (D-loop), and other elements present in *Malassezia* mitogenomes. The information contained in this file holds significance across various fields, including taxonomy, ecology, physiology, and evolution. The GFF3 file's annotation consists of nine columns, eachserving a specific purpose:

1. Identity: This column provides the name of the scaffold where the annotated sequenceis located.
2. Method: Here, the algorithm or operating procedure responsible for generating thesequence is described, such as “mitos” or “mitfi”.
3. Type: This column denotes the name of the annotated resource type, which could be
   “gene”, “exon”, “mRNA”, “tRNA”, “rRNA”or “ncRNA_gene”.
4. Start: Indicates the position from which the genomic resource is noted within thescaffold, with a displacement of one base.
5. End: Represents the end position of the annotated genomic resource, with an offset of
   one base.
6. Score: This value indicates the confidence level of the source in the annotated resource.When the confidence level is indeterminate, a dot “.” is used.
7. Strand: The presence of a single character indicates the strand of the annotated resource.“+” sign represents the positive or 5'->3' strand,“-” denotes the negative or 3'->5' strand,and “.” indicates an undetermined strand.
8. Phase: This column uses values 0, 1, or 2 to indicate the resource phase of the coding sequence (CDS), while “.” is used for other types of resources.
9. Attributes: This column contains all other relevant information related to the annotatedresource.

In addition to the consolidated GFF3 file, we are pleased to inform that we have also made available the annotations of the 45 mitogenomes in two additional formats: BED-6 and TBL. These annotations offer further insights into the genomic features of the *Malassezia* mitogenomes. The BED-6 format provides a concise representation of the annotation data andincludes the following six columns:

1. Chromosome/Identity: The name of the scaffold where the annotated sequence islocated.
2. Start: The position from which the genomic resource is noted within the scaffold, witha displacement of one base.
3. End: The end position of the annotated genomic resource, with an offset of one base.
4. Name: A label or identifier for the annotated resource.
5. Score: This value indicates the confidence level of the source in the annotated resource.When the confidence level is indeterminate, a dot "." is used.
6. Strand: The presence of a single character indicates the strand of the annotated resource.

   "+" sign represents the positive or 5'->3' strand, "-" denotes the negative or 3'->5' strand,and "." indicates an undetermined strand.

On the other hand, the TBL format provides a tab-separated table that includes more detailed information about the annotation data, enhancing the understanding of the genomic features. It encompasses various columns that capture relevant details about each annotated resource, such as start and end positions, strand orientation, attributes, and more.

We have also made available the data in the FASTA format. These FASTA files contain the complete mitochondrial genomes of the organisms, along with the nucleotide and protein sequences of protein-coding genes, tRNAs, rRNAs, and other elements present in the *Malassezia* mitogenomes.

The FASTA format is widely used in bioinformatics and genomics research, as itprovides a simple and efficient way to represent nucleotide and protein sequences. Eachsequence in the FASTA file is denoted by a header line, starting with a ">" symbol, followed by a description or identifier for the sequence. The actual sequence data follows the header line.The availability of the complete mitochondrial genomes in the FASTA format allows researchers to conduct in-depth analyses, such as sequence alignments, phylogenetic studies,and comparative genomics. Moreover, the individual sequences of protein-coding genes,tRNAs, rRNAs, and other elements offer valuable insights into the functional and structural aspects of the *Malassezia* mitogenomes.

Implementation of *Malassezia*DB

To ensure easy access and utilization of the valuable dataset, we established a robust infrastructure using the Machado Framework (Mudadu and Zerlotini, 2020) for the database. The entire dataset, along with necessary scripts and a Docker Compose file, is available for retrieval from the Open Science Framework (OSF) repository.

Efficient population of the database tables is facilitated by a custom script that handlesdata from FASTA and GFF files. This script is provided in the OSF repository, allowing researchers to query and analyze genomic information effortlessly. To support local executionof the database, we included a Docker Compose file in the OSF repository. Researchers can easily initiate an instance of the database and perform in-depth analyses on *Malassezia* mitochondrial genomes without complex setup procedures.

The user-friendly interface of the Mitogenome Database (Figure 2) enables users to interact with and explore available mitogenomes with intuitive navigation and search functionalities.

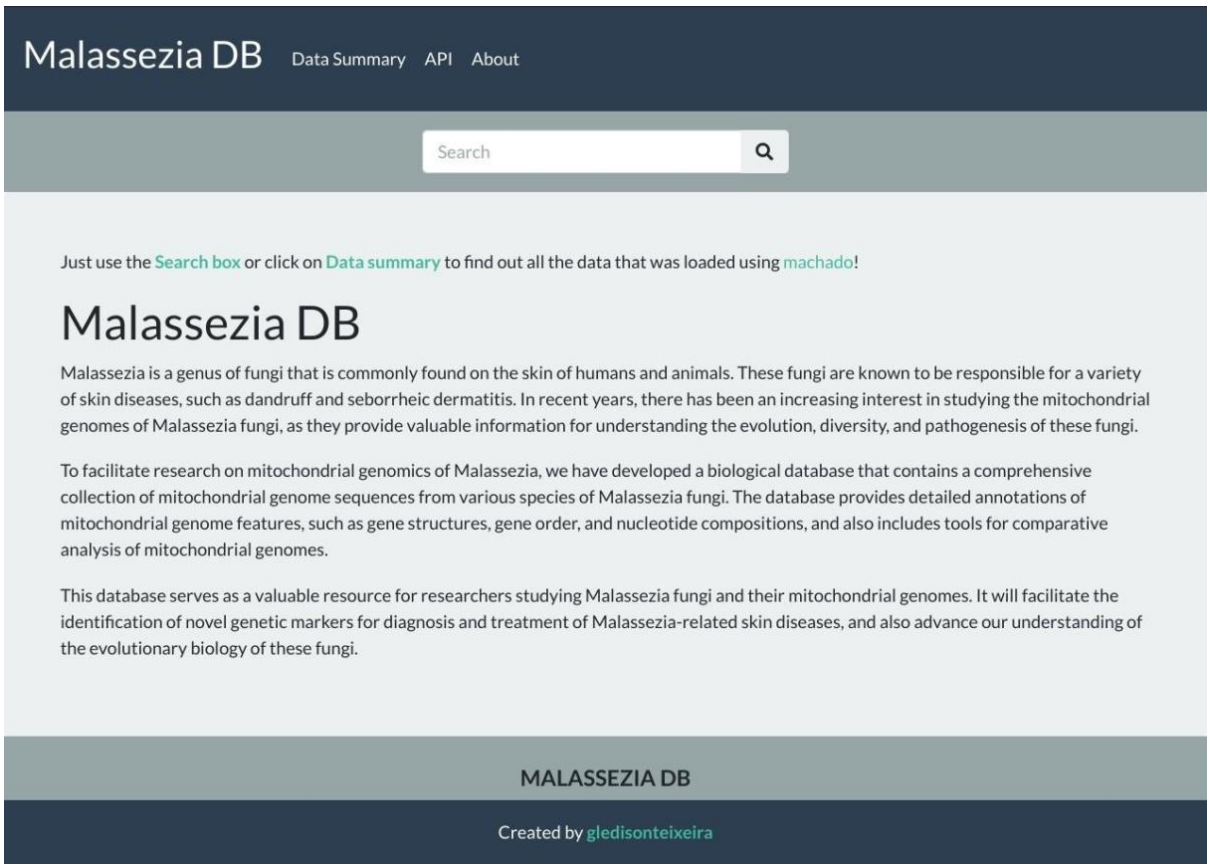


Figure 2. Interface of the Mitogenome Database

The research and display interface of the database (Figure 3) allows users to execute sophisticated queries, visualize mitochondrial genomes, and access detailed annotation information, providing valuable insights into genomic features.

The Genome Browser is a complementary tool that enables researchers to visualize the positional information of genes and genetic elements within *Malassezia* mitochondrial genomesin a graphical format. It offers an interactive and dynamic display of annotated features, allowing users to explore genomic organization and arrangement easily.

The Genome Browser enhances the exploration and interpretation of genetic elements,helping researchers identify potential regulatory regions, conserved motifs, and structural variations. It aids in the identification of gene clusters and overlaps, contributing to functional relationship and evolutionary insights.

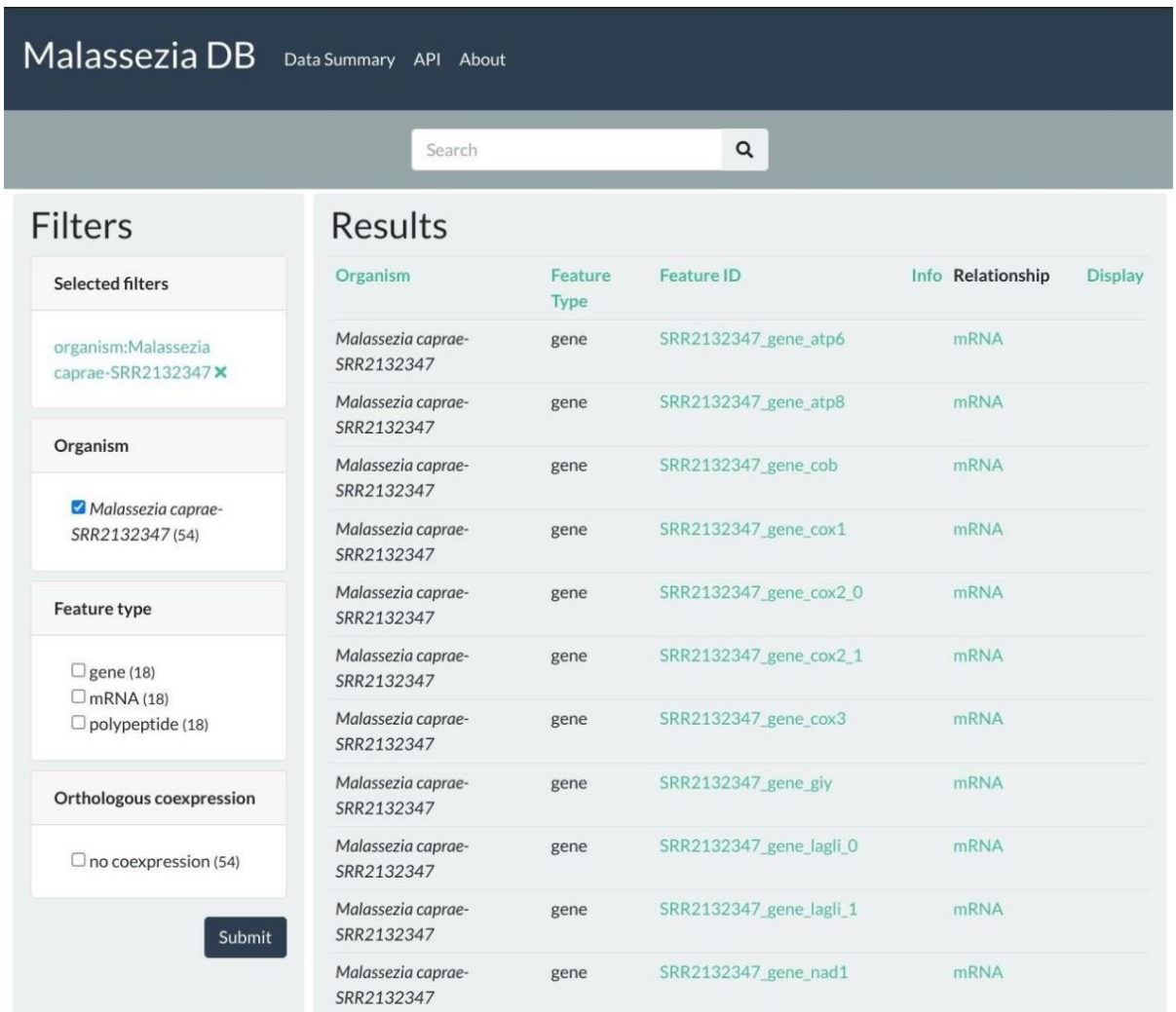


Figure 3. Search and display interface of the Database

Figure 4 provides a glimpse of the Genome Browser interface, showcasing its capability to empower researchers in understanding the genomic landscape of *Malassezia* organisms.

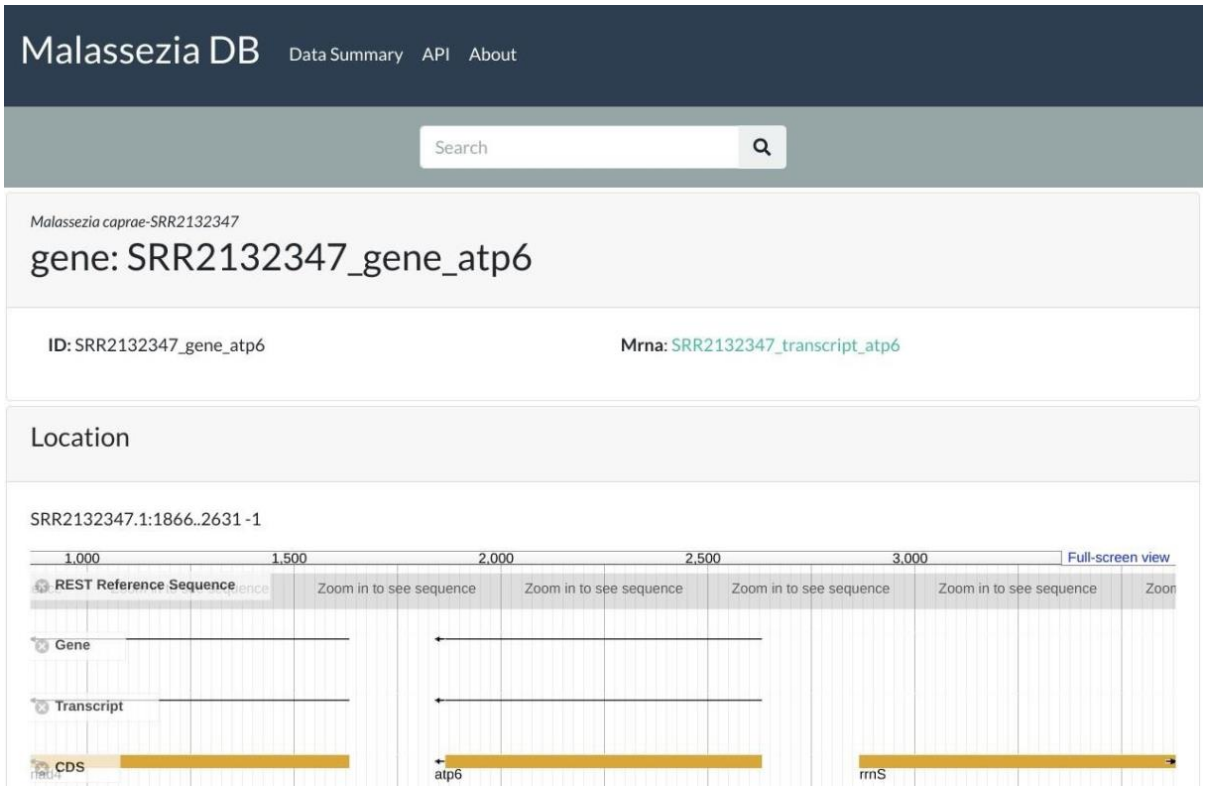


Figure 4. Genome Browser

The Chado schema implemented in *Malassezia*DB (Figure 5) serves as the backbone for organizing and structuring genomic data. This well-established and widely used relational database schema ensures efficient and coherent data retrieval and management, fostering collaboration among researchers.

The Chado schema's robustness and versatility within *Malassezia*DB support advanced genomic analyses and interpretation, driving advancements in fungal genomics and facilitating discoveries that benefit multiple scientific disciplines.

Overall, the integrated *Malassezia*DB platform presents a powerful resource for researchers to conduct comprehensive and visually intuitive investigations into the mitochondrial genomes of *Malassezia* organisms. We encourage users to leverage these tools and resources for furthering their research endeavors in the field of fungal genomics.

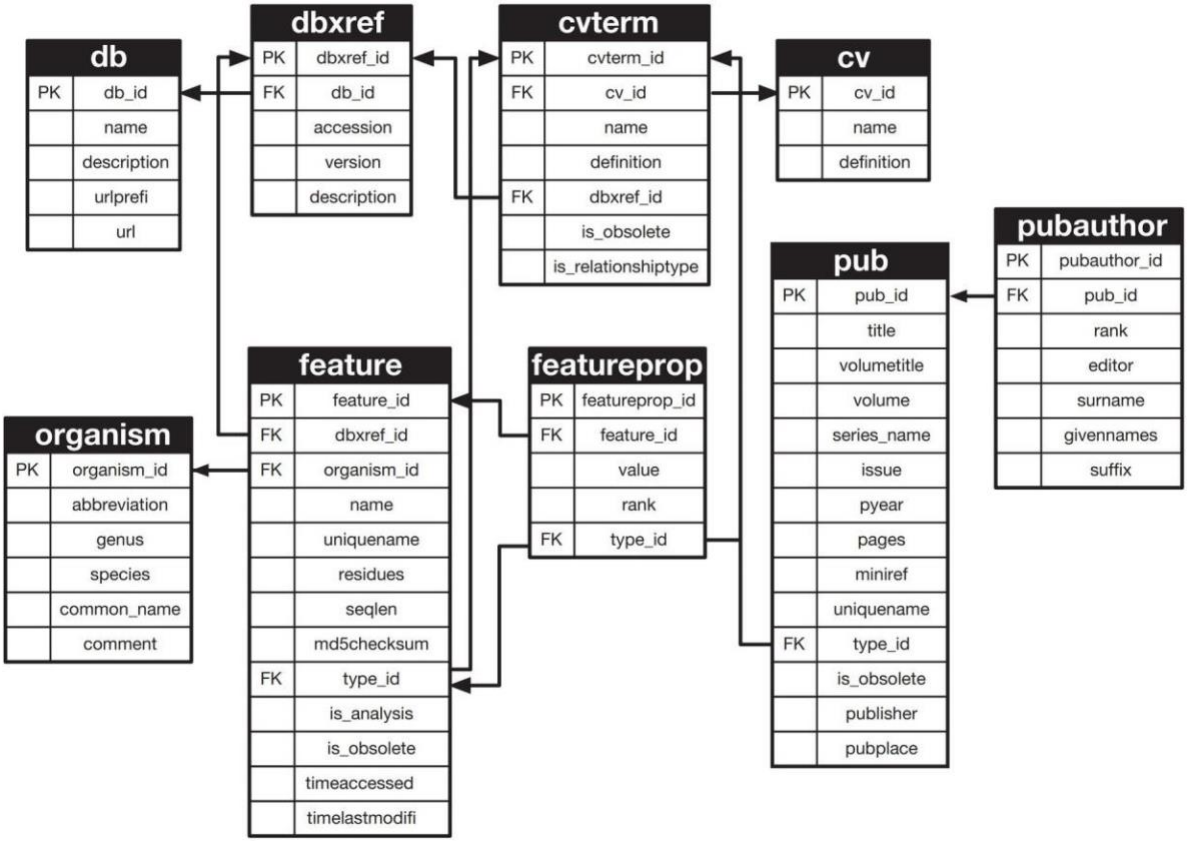


Figure 5. Chado Schema

Implementation of *Malassezia*BLAST

*Malassezia*BLAST offers researchers the unique capability to conduct comprehensive comparisons and analyses of gene and protein sequences derived from the 45 mitochondrial genomes of 10 distinct *Malassezia* species, all of which are available in the comprehensive *Malassezia* Dataset and *Malassezia*DB. Seamlessly integrated with SequenceServer (Priyam *et al.*, 2019), this cutting-edge platform facilitates sequence alignment and exploration, providing valuable insights into the genetic composition of these fungal organisms.

To initiate sequence alignment, users can conveniently input gene sequences or import sequence files directly into the designated input box (Figure 6). For a personalized approach, researchers can fine-tune parameters, such as -e-value, -num_alignments, and more, through theoption box, ensuring tailored and precise results. The BLAST output delivers comprehensive alignment results, including the queried sequence and corresponding

references from the database, facilitating the identification of homologous genes or specificgenes of interest (Figure 7).



Figure 6. Interface of the *Malassezia*BLAST

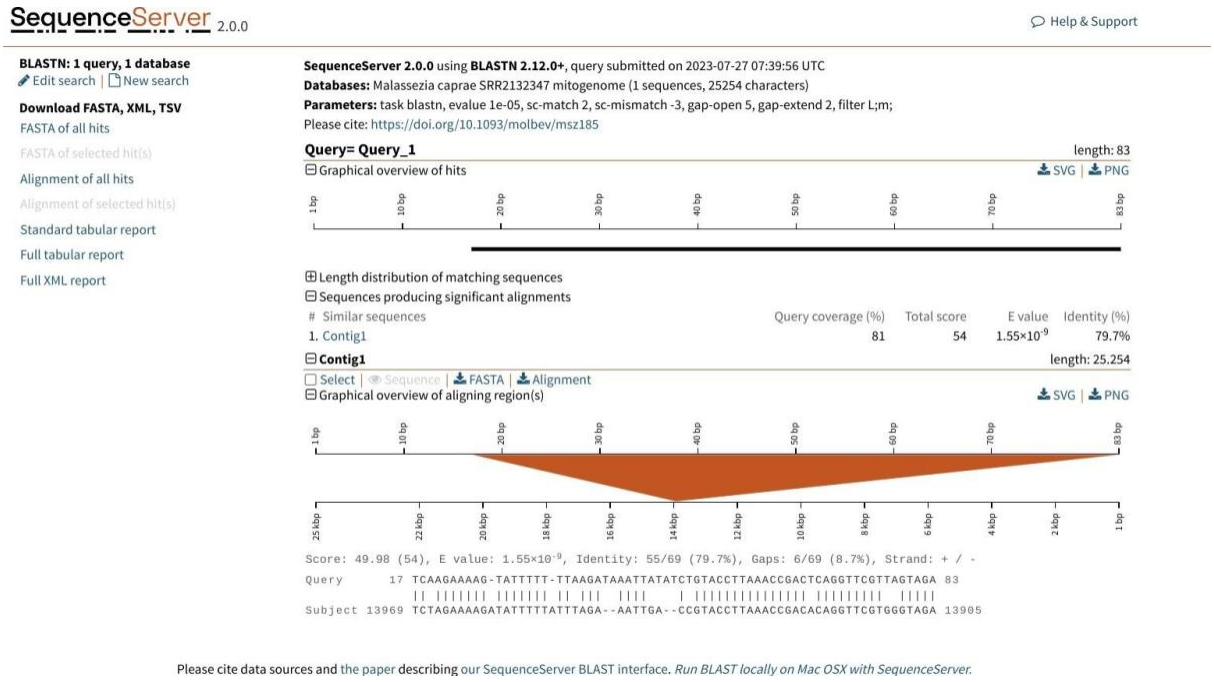


Figure 7. Similarity search result in MalasseziaBLAST interface

Subject area: Biological Sciences; Microbiology

Data Access

Inform where the data can be accessed:

( ) Together with the article: when the data are submitted with the article and will beaccessible in LADS itself.

(x) In a qualified institutional repository: LADS suggests depositing in Mendeley Data. Otherrepositories will be evaluated.

In case of a repository, please inform

Repository name: Open Science Framework (OSF)

DOI of the dataset (when available):

https://doi.org/10.17605/OSF.IO/QMZKSLink to access the data:

https://osf.io/qmzks/

Reference

1. Al Arab M, Zu Siederdissen CH, Tout K, Sahyoun AH, Stadler PF, Bernt M. Accurateannotation of protein-coding genes in mitochondrial genomes. Molecular phylogeneticsand evolution. 2017;106: 209-216.
2. Altschul, S. F., *et al.* Basic Local Alignment Search Tool. Journal of Molecular Biology215, 403–410 (1990) doi:10.1016/S0022-2836(05)80360-2
3. Aykut B, Pushalkar S, Chen R *et al.* The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. Nature. 2019;574:264–7.
4. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, Guo E, Dider S, Zhang J, Bridge C, Hogue G, Duncan A, Morgan M, Flores T, Bimber BN, Haw R, Cain S, Buels RM, Stein LD, Holmes IH. JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol. 2023 Apr 17;24(1):74. doi: 10.1186/s13059-023-02914-z. PMID: 37069644; PMCID: PMC10108523.
5. Findley K, Oh J, Yang J *et al.* Topographic diversity of fungal and bacterial communities in human skin. Nature. 2020;498:367–70.
6. Huszcz G.B; Teixeira G.E; Rodrigues-Oliveira I.H; Rocha R.R; Oliveira RC; Takao PS; Aciole Barbosa D.; Jabes DL; Kavalco KF; Pasa R; Menegidio F.B. Five mitochondrial genomes of the lipophilic fungus *Malassezia* (*Malassezia*les, Exobasidiomycetes). RevistaFT, v. 27, n. 124, p. 1-2, 2023.
7. Limon JJ, Tang J, Li D *et al. Malassezia* is associated with Crohn's disease and exacerbates colitis in mouse models. Cell Host Microbe. 2019;25:377–88.
8. Mudadu MA, Zerlotini A. Machado: Open source genomics data integration framework. Gigascience. 2020 Sep 14;9(9):giaa097. doi:10.1093/gigascience/giaa097. PMID: 32930331; PMCID: PMC7490629.
9. Mungall C.J., Emmert D.B., FlyBase Consortium. (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. Bioinformatics, 23, i337–i346.
10. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, Chowdhary H, Pieniak I, Maynard LJ, Gibbins MA, Moon H, Davis-Richardson A, Uludag M,Watson-Haigh NS, Challis R, Nakamura H, Favreau E, Gómez EA, Pluskal T, Leonard G, Rumpf W, Wurm Y. Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. Mol Biol Evol. 2019 Dec 1;36(12):2922-2924. doi: 10.1093/molbev/msz185. PMID: 31411700; PMCID: PMC6878946.
11. Saunte DML, Gaitanis G, Hay RJ. *Malassezia*-Associated skin diseases, the use of diagnostics and treatment. Front Cell Infect Microbiol. 2020;10:112.
12. Spatz M, Richard ML. Overview of the potential role of *Malassezia* in gut health and disease. Front Cell Infect Microbiol. 2020;10:201.
13. Theelen B, Cafarchia C, Gaitanis G *et al. Malassezia* ecology, pathophysiology, and treatment [published correction appears in Med Mycol. 2019 Apr 1;57(3):e2]. Med Mycol. 2018;56:S10–25.

6. DISCUSSÃO

A presente dissertação teve como objetivo explorar a diversidade mitogenômica do gênero *Malassezia*, um grupo de fungos com ampla relevância ecológica e taxonômica. Como propósito de avançar no entendimento desses organismos, foi realizado um amplo e detalhado conjunto de análises, culminando na construção de um *dataset* mitogenômico abrangente e na implementação de um banco de dados relacional para armazenar e disponibilizar as informações obtidas.

Inicialmente, foram prospectados genomas mitocondriais completos disponíveis em bases de dados públicas, bem como bibliotecas de sequenciamento de nova geração de genoma completo (*WGS*) de espécies de *Malassezia* ainda não representadas nos mitogenomas conhecidos. Nesse sentido, apresentamos nessa dissertação, e no artigo Huszcz et al. (2023), cinco novos mitogenomas, compreendendo as espécies *Malassezia caprae*, *Malassezia equina*, *Malassezia dermatis*, *Malassezia nana* e *Malassezia vespertilionis*. Esses se somaram aos 192 mitogenomas de basidiomicetos, entre os 1.189 genomas mitocondriais de fungos disponíveis na base de dados Organelle Genome Resources.

Como esperado, todos os mitogenomas de *Malassezia* apresentaram 14 genes codificadores de proteínas, que estão presentes em quase todos os fungos e metazoários, além do gene rps3, que está presente apenas em 65% dos basidiomicetos (Korovesi et al., 2018). Da mesma forma, os mitogenomas de *Malassezia* analisados demonstraram uma variação substancial entre as espécies, resultante da variação do tamanho, frequência do íntron e regiões intergênicas, levando a tamanhos de mitogenoma variando de ~20-50kb. No entanto, em fungos, essa variação na ordem dos genes é conhecida e observada (Zhao et al., 2013; Aguileta et al., 2014; Moreno-Indias et al., 2021). Mesmo considerando apenas o gênero *Malassezia*, essa diversidade na organização do genoma mitocondrial foi observada durante o presente trabalho.

Os potenciais genes codificadores (PCGs) das espécies de *Malassezia* incluíram 14 PCGs relacionados à fosforilação oxidativa [NAD desidrogenase (nad1-6 e nad4L), citocromo oxidases (cox1, cox2 e cox3), citocromo b (cob) e 3 ATP sintases (atp6 , atp8 e atp9), proteína ribossomal S3 (rps3)] e 2 genes de RNA ribossomal (rns e rnl). Essa distribuição foi consistente com a observada em mitogenomas típicos de fungos, como os pertencentes a *Ascomycota* e *Basidiomycota*. Nesse sentido, a organização gênica parece ser mais discrepante entre as espécies basais e tende a se estabilizar nos clados

filogenéticos recentes.

Em geral, as análises do mitogenoma fúngico apresentam considerável variação no tamanho do genoma, o que até agora tem contribuído para suas alterações de sequência decorrentes de mutação e recombinação (Pantou et al., 2008; Christinaki et al., 2022). Por outrolado, em plantas e mais especificamente em genomas de cloroplastos, a diversidade de tamanho pode ser explicada pela complexidade da sequência, bem como pela quantidade de DNA repetitivo (Palmer, 1990). Os genomas de cloroplastos (cpDNA) geralmente apresentam uma estrutura comumente encontrada de duas regiões únicas divididas por repetições invertidas longas (long inverted repeats, LIRs) (Jansen et al., 2005) permitindo que o cpDNA evolua sob forte restrição - seja mecanicista ou seletiva - para manter um genoma compacto e amplamente gênico (Palmer, 1990). Em fungos, a falta de LIRs na maioria das espécies indica uma evolução livre dos mitogenomas em relação ao seu tamanho, com outros mecanismos subjacentes propostos (Hao, 2022).

Baseado nos resultados apresentados em Huszcz et al. (2023), a montagem e anotação desses novos genomas mitocondriais foram conduzidas com sucesso, permitindo a comparação filogenética com outros fungos do mesmo grupo taxonômico. Além disso, essa análise revelou *insights* valiosos sobre as relações evolutivas e a diversidade genômica dentro do gênero.

A organização cuidadosa dos dados também resultou na construção de um *dataset* curado, contendo informações relevantes sobre a genômica estrutural e funcional dos mitogenomas identificados e montados. Esse *dataset* representa uma valiosa fonte de conhecimento e está disponível para a comunidade científica, possibilitando pesquisas futuras e reutilização dos dados para diferentes propósitos.

Foi demonstrada a importância da implementação de um banco de dados relacional para armazenar informações sobre genomas mitocondriais e suas anotações funcionais. Nesse sentido, o MalasseziaDB foi pioneiramente desenvolvido como uma solução eficaz, permitindo o acesso contínuo e organizado dos dados obtidos. Ele oferece aos usuários a opção de instalar espelhos totalmente funcionais do banco em seus próprios ambientes de trabalho ou em plataformas de IaaS/PaaS. O MalasseziaDB representa uma prova de conceito funcional para o desenvolvimento de uma estrutura de banco de dados biológicos baseada no conceito de DbaaS. É notável por ser a primeira aplicação a utilizar contêineres Docker e o framework Machado para armazenar genomas mitocondriais completos e anotações funcionais. Além disso, destaca-se por ser o primeiro projeto a disponibilizar um banco de dados focado em mitogenômica, permitindo que os usuários executem o banco em sua própria infraestrutura computacional. Vale ressaltar que, no

contexto mencionado, apenas o ParaDB (Aciole Barbosa et al., 2019) oferece uma estrutura semelhante, fornecendo um serviço deDbaaS voltado para a genômica funcional dos fungos patogênicos do gênero *Paracoccidioides*.

Ao comparar o Machado e o MalasseziaDB com projetos similares, podemos observar que o tripal (Ficklin et al., 2011; Sanderson et al., 2013; Spoor et al., 2018) é a aplicação mais semelhante disponível. O Tripal é um projeto bem estabelecido e amplamente apoiado por umaextensa comunidade. Ele oferece uma plataforma de banco de dados de código aberto que combina o esquema de banco de dados GMOD Chado (Mungall et al., 2007) com o Drupal, umsistema de gerenciamento de conteúdo usado por 1,7% de todos os sites da internet, correspondendo a 3% do mercado global de sistemas de gerenciamento de conteúdo (W3techs, 2019). As diferenças de uso entre esses sistemas de gerenciamento de conteúdo também se refletem no tamanho de suas comunidades de desenvolvimento e suporte, bem como na quantidade de plugins, widgets e templates disponíveis para cada um deles. Assim como o machado, o tripal é totalmente integrado ao esquema gmod chado, fornecendo diversas estruturas de dados para armazenar informações sobre organismos, bibliotecas genômicas, estoques biológicos, coleções e recursos genômicos, além de seus alinhamentos, anotações e outras informações genômicas (Ficklin et al., 2011; Sanderson et al., 2013; Spoor et al., 2018). Além disso, o projeto Tripal oferece módulos de extensão que permitem o carregamento de dados a partir de diferentes formatos de arquivos, como FASTA,GFF, OBO, GAF, BLAST XML, KEGG e InterProScan XML (Ficklin et al., 2011; Sanderson et al., 2013; Spoor et al., 2018). Essa versatilidade do Tripal, graças a sua variedade de estruturas de dados e módulos, também torna o sistema complexo em sua implementação, especialmente devido à necessidade de lidar com diferentes formatos dedados que podem não ser reconhecidos inicialmente pelos scripts da aplicação. Além disso, o uso de conceitos de taxonomias provenientes da estrutura do Drupal e uma maior necessidade de codificação podem aumentar a complexidade. Nesse sentido, a escolha do Machado se mostra eficaz para reduzir a complexidade na implementação do MalasseziaDB. Por fim, a criação de uma aplicação web para pesquisas *BLAST* nas sequências mitogenômicas de *Malassezia* agrega uma ferramenta poderosa para identificar sequências homólogas em diferentes organismos. Isso pode abrir novas perspectivas de estudo,apontando para possíveis funções e relações funcionais ainda não exploradas.

7. CONSIDERAÇÕES FINAIS

Ao finalizar esta dissertação, evidenciamos a importância de estudar a diversidade mitogenômica de fungos do gênero *Malassezia*, bem como as potencialidades proporcionadas pela integração de técnicas de bioinformática e análises filogenéticas. Os resultados alcançadosnão apenas contribuem para o conhecimento específico sobre esses fungos, mas também trazemavanços metodológicos que podem ser aplicados em outras pesquisas relacionadas à genômicafúngica.

Dessa forma, esperamos que os esforços empreendidos neste estudo estimulem novas investigações e colaborações em um campo de pesquisa cada vez mais relevante. Ao disponibilizar os dados e ferramentas desenvolvidos, almejamos contribuir para o avanço da ciência e promover uma maior compreensão da diversidade e evolução dos fungos do gênero *Malassezia* e sua importância nos ecossistemas e interações com outros organismos. Acreditamos que esses esforços coletivos são essenciais para enfrentar os desafios futuros no campo da biologia molecular e genômica, e para desvendar os segredos ainda desconhecidos do mundo fascinante dos fungos.

8. PERSPECTIVAS FUTURAS

Esta dissertação abriu portas para novas perspectivas emocionantes na pesquisa da mitogenômica de fungos do gênero *Malassezia*. Com base nos resultados e nas ferramentas desenvolvidas, algumas direções promissoras podem ser exploradas para avançar ainda mais em nosso entendimento desses organismos fascinantes e sua relevância em diversos contextos biológicos e clínicos:

1. Exploração de Espécies Adicionais: Com o estabelecimento de um fluxo de trabalho eficiente para a montagem e anotação de mitogenomas, a busca por novos genomas mitocondriais de *Malassezia* pode ser ampliada. A inclusão de espécies adicionais dentro do gênero pode fornecer uma compreensão mais abrangente das adaptações genômicas e das variações funcionais presentes em diferentes linhagens.
2. Análises Funcionais e Evolutivas: Além das análises filogenéticas comparativas, a exploração das anotações funcionais dos mitogenomas pode revelar informações preciosas sobre a evolução e as funções biológicas específicas de genes e elementos genéticos presentes em *Malassezia*. A investigação de genes codificadores de proteínase elementos não-codificadores pode abrir caminho para descobertas sobre vias metabólicas, adaptações e interações ecológicas.
3. Estudo da Expressão Gênica Mitocondrial: A compreensão da expressão gênica mitocondrial em diferentes estágios de desenvolvimento ou sob diferentes condições ambientais pode proporcionar *insights* sobre a regulação da atividade mitocondrial em *Malassezia*. Análises de transcriptômica mitocondrial podem revelar padrões de expressão gênica relevantes para a fisiologia e patogenicidade desses fungos.
4. Investigação de Interações Hospedeiro-Parasita: Dado o papel conhecido de algumas espécies de *Malassezia* como comensais ou patógenos oportunistas em humanos e animais, investigações futuras podem se concentrar na interação entre esses fungos e seus hospedeiros. Estudos que abordem a expressão gênica em ambientes hospedeiros podem fornecer *insights* sobre os mecanismos devirulência e resposta imune associadosa infecções por *Malassezia*.

5. Integração de Metagenômica e Mitogenômica: Em ambientes complexos, como a microbiota da pele ou do couro cabeludo, a integração de dados metagenômicos com análises mitogenômicas pode aprimorar nossa compreensão das interações

ecológicas de *Malassezia* com outros microrganismos e seus impactos no ecossistema hospedeiro.

6. Aprofundamento das Ferramentas e Recursos: A aplicação web para pesquisas *BLAST* nas sequências mitogenômicas de *Malassezia* pode ser continuamente aprimorada, coma inclusão de recursos adicionais e funcionalidades para melhorar a análise e a interpretação dos resultados. A implementação de uma interface interativa e amigável pode tornar a ferramenta ainda mais acessível e útil para a comunidade científica.
7. Expansão da Comunidade Científica: O compartilhamento de dados e recursos, bem como a colaboração entre diferentes grupos de pesquisa, são fundamentais para impulsionar ainda mais o campo da mitogenômica de *Malassezia*. A formação de uma comunidade científica engajada e colaborativa pode acelerar a descoberta de novos conhecimentos e estabelecer estratégias conjuntas para enfrentar desafios científicos emergentes.

Essas perspectivas futuras representam apenas algumas das muitas possibilidades que a mitogenômica de *Malassezia* pode oferecer. Acredita-se que a continuidade do estudo dessesfascinantes fungos pode contribuir significativamente para a compreensão da biodiversidade fúngica, a evolução genômica e a relação complexa entre esses microrganismos e seus hospedeiros. Com a dedicação contínua da comunidade científica e o avanço das tecnologias de sequenciamento e análise de dados, novos horizontes serão desvendados, revelando um mundo ainda mais intrigante e surpreendente no reino dos fungos do gênero *Malassezia*.

9. REFERENCIAS

Abdillah A, Ranque S. Chronic Diseases Associated with *Malassezia* Yeast. J Fungi (Basel) [internet] 2021 [cited 2023 oct 12]; 7(10):855. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8540640/pdf/jof-07-00855.pdf. doi: 10.3390/jof7100855. Pubmed PMID: 34682276; Pubmed Central PMCID: PMC8540640.

Aciole Barbosa D, Menegidio FB, Alencar VC, Gonçalves RS, Silva JFS, Vilas Boas RO, et al. ParaDB: Amanually curated database containing genomic annotation for the human pathogenicfungi Paracoccidioides spp. PLoS Negl Trop Dis. [internet] 2019 [cited jul 15];13(7):e0007576. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6658007/pdf/pntd.0007576.pdf. doi: 10.1371/journal.pntd.0007576. Pubmed PMID: 31306428; Pubmed Central PMCID: PMC6658007.

Adalsteinsson JA, Kaushik S, Muzumdar S, Guttman-Yassky E, Ungar J. An update on the microbiology, immunology and genetics of seborrheic dermatitis. Exp Dermatol. [internet] 2020 [cited may 12];29(5):481-489. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1111/exd.14091. doi: 10.1111/exd.14091.

Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M et al. The Galaxyplatform for accessible, reproducible and collaborative biomedical analyses: 2018update. Nucleic acids research. 2018;46(W1):W537-W544.

Aguileta, G. et al. High variability of mitochondrial gene order among fungi. GenomeBiol. Evol. 2014;6(2):451–465.

Al Arab M, Zu Siederdissen CH, Tout K, Sahyoun AH, Stadler PF, Bernt M. Accurate annotation of protein-coding genes in mitochondrial genomes. Molecular phyloge-netics and evolution. 2017;106: 209-216.

Alikhan, NF., Petty, N.K., Ben Zakour, N.L. et al. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. [internet] 2011 [cited may 12];12: 402.Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3163573/pdf/1471-2164-12-402.pdf. doi: https://doi.org/10.1186/1471-2164-12-402

Altschul, S. F. et al. Basic Local Alignment Search Tool. Journal of Molecular Bio-logy [internet] 1990 [cited jul 15];215:403–410. Available from: https://www.sciencedirect.com/science/article/pii/S0022283605803602?via%3Dihub doi:10.1016/S0022-2836(05)80360-2.

Ambrosio, A. B., do Nascimento, L. C., Oliveira, B. V. L., Teixeira, P. J. P., Tiburcio, R. A. et al. Global analyses of Ceratocystis caca-ofunesta mitochondria: from genome to proteome. BMC Genomics [internet] 2013 [cited jul 15];14:913. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3605234/pdf/1471-2164-14-91.pdf Pubmed PMID: 23394930; Pubmed Central PMCID: PMC3605234

Araújo D. S., De-Paula R. B., Tomé L. M. R., Quintanilha-Peixoto G., Salvador-Mon- toya C. A., Del-Bem L.-E. et al. Comparative mitogenomics of Agaricomy- cetes: diversity,

abundance, impact and coding potential of putative open-reading fra-mes. Mitochondrion 2021 May;58:1–13.

Ashbee, H. R., Ingham, E., Holland, K. T., and Cunliffe, W. J. The carriage of *Malassezia* furfur serovars A, B and C in patients with pityriasis versicolor, seborr- hoeic dermatitis and controls. Br. J. Dermatol. [internet] 2013 [cited jul 15];129:533–540. Available from: https://pubmed.ncbi.nlm.nih.gov/8251348/. doi:10.1111/j.1365-2133.1993.tb00480.

Ashbee HR, Evans EG. Immunology of diseases associated with *Malassezia* species. Clin Microbiol Rev. [internet] 2002 Jan[cited jul 15];15(1):21-57. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC118058/pdf/cm0002.pdf. doi: 10.1128/CMR.15.1.21-57.2002. PMID: 11781265; Pubmed Central PMCID: PMC118058.

Avin F. A., Subha B., Tan Y.-S., Braukmann T. W. A., Vikineswary S., Hebert P. D. N. Escaping introns in COI through cDNA barcoding of mushrooms: pleurotus as a test case. Ecol. Evol. [internet] 2017 [cited jul 15];7:6972–6980. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5587487/pdf/ECE3-7-6972.pdf. doi: 10.1002/ece3.3049. Pubmed PMID: 28904776; Pudmed Central PMCID: PMC5587487.

Aykut B, Pushalkar S, Chen R et al. The fungal mycobiome promotes pancreatic onco-genesis via activation of MBL. Nature. 2019;574:264–7.

Baillon H. Traité de botanique médicale cryptogamique.Paris: Octave Doin; 1889. p234–235.

Balaji, H., Heratizadeh, A., Wichmann, K., Niebuhr, M., Crameri, R., Scheynius, A. et al. *Malassezia* sympodialis thioredoxin-specific T cells are highly cross-reac-tive to human thioredoxin in atopic dermatitis. J. Allergy Clin. Immunol. [internet] 2011 [cited jul 15];128, 92– 99 e94. Available from: https://www.sciencedirect.com/science/article/pii/S0091674911003770?via%3Dihub. doi: 10.1016/j.jaci.2011.02.043

Batra R, Boekhout T, Guého E, Cabañes FJ, Dawson TL Jr, Gupta AK. Malassezia Bail-lon, emerging clinical yeasts. FEMS Yeast Res. [internet] 2005 Dec [cited May 12];5(12):1101-13. Available from: https://www.sciencedirect.com/science/article/pii/S0091674911003770?via%3Dihub. doi: 10.1016/j.femsyr.2005.05.006. Pubmed PMID: 16084129.

Bilska, K., Kulik, T., Ostrowska-Kołodziejczak, A., Buśko, M., Pasquali, M., Beyer, M. et al. Development of a highly sensitive FcMito qPCR assay for the quantifi-cation of the toxigenic fungal plant pathogen Fusarium culmorum. Toxins [internet] 2018 [cited Jul 15];10:211. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5983267/pdf/toxins-10-00211.pdf. doi: 10.3390/toxins10050211.

Bizzozero J. Über die Mykrophyten der normalen Oberhaut des Menschen. Vir-chow Arch Pathol Anat Physiol 1884;98:441-459.

Bjelland T, Ekman S. Fungal diversity in rock beneath a crustose lichen as revea-led by molecular markers. Microbiol Ecol 2005;49:598–603.

Bjerre, R. D., Bandier, J., Skov, L., Engstrand, L., and Johansen, J. D. The roleof the skin

microbiome in atopic dermatitis: a systematic review. Br. J. Dermatol. [internet] 2017 [cited Jul 15];177:1272–1278. Available from: https://onlinelibrary.wiley.com/doi/10.1111/ddg.14709. doi: 10.1111/bjd.15390.

Boekhout T, Renting M, Scheffers WA et al. The use of karyotyping in the syste-matics of yeasts. Antonie van Leeuwenhoek 1993;63:157-163.

Bolender, N., Sickmann, A., Wagner, R., Meisinger, C., Pfanner, N. Multi-ple pathways for sorting mitochondrial precursor proteins. EMBO Rep. [internet] 2008 [cited Jul 15]; 9:42–49. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2246611/pdf/7401126.pdf. doi: 10.1038/sj.embor.7401126.

Borda, L. J., Wikramanayake, T. C. Seborrheic Dermatitis and Dandruff: AComprehensive Review. J. Clin. Invest. Dermatol. [internet] 2015 [cited Jul 15];3, 1-10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4852869/pdf/nihms754376.pdf. doi: 10.13188/2373- 1044.1000019.

Brodska, P., Panzner, P., Pizinger, K., and Schmid-Grendelmeier, P. IgE-media-ted sensitization to *Malassezia* in atopic dermatitis: more common in male patients and in head and neck type. Dermatitis [internet] 2014 [cited Jul 15];25, 120–126. Available from: https://www.zora.uzh.ch/id/eprint/108690/1/Brodska%20et%20al.%20-%20Dermatitis%202014.pdf. doi: 10.1097/ DER.0000000000000040.

Bullerwell, C. E., Lang, B. F. Fungal evolution: the case of the vanishing mitochondrion. Curr. Opin. Microbiol. [internet] 2005 [cited Jul 15];8:362–369. Available from: https://www.sciencedirect.com/science/article/pii/S1369527405000834?via%3Dihub. doi: 10.1016/j.mib.2005.06.009

Carvalho D. S., Andrade R. F. S., Pinho S. T. R., Góes-Neto A., Lobão T. C. P., Bom- fim G. C. et al.What are the evolutionary origins of mitochondria? A com- plex network approach. PLoS One [internet] 2015 Sep 2 [cited Jul 15] 10(9):e0134988. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4557972/pdf/pone.0134988.pdf. doi: 10.1371/journal.pone.0134988

Castellani A, Chalmers AJ. Manual of tropical medicine. 2ª ed. London: Baillière, Tindal and Cox; 1913. 1747 p.

Chng, K. R., Tay, A. S., Li, C., Ng, A. H., Wang, J., Suri, B. K. et al. Wholemetagenome profiling reveals skin microbiome-dependent susceptibility to atopicdermatitis flare. Nat. Microbiol. [internet] 2016 [cited Jul 15];1:16106. Available from: https://www.nature.com/articles/nmicrobiol2016106. doi: 10.1038/nmicrobiol.2016.106.

Crespo Erchiga V, Ojeda Martos A, Vera Casano A et al. Isolation and identifica-tion of Malassezia spp. in pityriasis versicolor, seborrheic dermatitis and healthy skin. Rev Iberoam Micol 1999;16:S16–S21.

Crespo Erchiga V, Guého E. Superficial diseases caused by Malassezia species. In: Merz WG, Hay RJ, editores. Topley and Wilson's microbiology and microbial infecti-ons, medical mycology. 10ª ed. London: Hodder Arnold, 2005. p. 202-19. v. 5.

Crespo-Erchiga, V., and Florencio, V. D. *Malassezia* yeasts and pityriasis versi-color. Curr. Opin. Infect. Dis. [internet] 2006 Apr. [cited Jul 15];19(2):139-147. Available from: https://pubmed.ncbi.nlm.nih.gov/16514338/. doi: 10.1097/01.qco.0000216624.21069.61. Pubmed PMID: 16514338.

Christinaki, A. C. et al. Mitogenomics and mitochondrial gene phylogeny decipher the evolution of Saccharomycotina yeasts. Genome Biol. Evol. 2002;14:evac073.

Dall'Oglio F, Nasca MR, Gerbino C, Micali G. An Overview of the Diagnosis and Management of Seborrheic Dermatitis. Clin Cosmet Investig Dermatol. [internet] 2022 Aug 6[cited Jul 15] ;15:1537-1548. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9365318/pdf/ccid-15-1537.pdf. doi: 10.2147/CCID.S284671. Pubmed PMID: 35967915; Pubmed Central PMCID: PMC9365318.

De Chiara, M., Friedrich, A., Barré, B., Breitenbach, M., Schacherer, J., Liti, G. Discordant evolution of mitochondrial and nuclear yeast genomes at popula-tion level. BMC Biol. [internet] 2020 [cited Jul 15];18:49. Available from: https://pubmed.ncbi.nlm.nih.gov/32393264/. doi: 10.1186/s12915-020-00786-4.

Deangelis, Y. M., Gemmer, C. M., Kaczvinsky, J. R., Kenneally, D. C., Schwartz, J. R., Dawson, T. L.Jr. Three etiologic facets of dandruff and seborrheic der- matitis: *Malassezia* fungi, sebaceous lipids, and individual sensitivity. J. Invest.Dermatol. Symp. Proc. [internet] 2005 [cited Jul 15]; 10(3): 295–297. Available from: https://pubmed.ncbi.nlm.nih.gov/16382685/. doi: 10.1111/j.1087-0024.2005.10119.

Deesi Z. The ecology and pathogenicity of the lipophilic yeasts (Malassezia spe-cies) – a study in Dubaï, United Arab Emirates. PhD Thesis, University of London, London, 2006.

Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle ge-nomes from whole genome data. Nucleic acids research. 2017;45(4):e18-e18.

Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A. et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol. [internet] 2023 Apr 17[cited Jul 15];24(1):74. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10108523/pdf/13059_2023_Article_2914.pdf. doi: 10.1186/s13059-023-02914-z. Pubmed PMID: 37069644; Pubmed Central PMCID: PMC10108523.

Dodge CW. Medical mycology: fungous diseases of men and other mammals. London: Forgotten Books;1935.

Mosby, St Louis Eichstedt E. Pilzbildung in der Pityriasis versicolor. Froriep Neue Notiz NaturHeilk 1846;39:270.

Fell JW, Scorzetti G, Connell L et al. Biodiversity of micro-eukaryotes in Antarc-tic Dry Valley soils with <5% soil moisture. Soil Biol Biochem 2006;38:3107–3119.

Findley, K., Oh, J., Yang, J., Conlan, S., Deming, C., Meyer, J. A., et al. Topo-graphic diversity of fungal and bacterial communities in human skin. Nature [internet] 2013 [cited Jul 15];498:367-370. Available from: https://www.nature.com/articles/nature12171. doi:

10.1038/nature12171.

Findley K, Oh J, Yang J et al. Topographic diversity of fungal and bacterial communities in human skin. Nature. 2020;498:367–70.

Fonseca P. L. C., Badotti F., De-Paula R. B., Araújo D. S., Bortolini D. E., Del-Bem L.-E. et al. Exploring the relationship among divergence time and coding and non-coding elements in the shaping of fungal mitochondrial genomes. Front. Micro-biol. [internet] 2020 Apr.29 [cited Jul 15];11:765. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7202290/pdf/fmicb-11-00765.pdf. doi: 10.3389/fmicb.2020.00765.

Formighieri E. F., Tiburcio R. A., Armas E. D., Medrano F. J., Shimo H., Carels N. et al. The mitochondrial genome of the phytopathogenic basidiomycete Monili-ophthora perniciosa is 109kb in size and contains a stable integrated plasmid. Mycol.Res. 2008;112:1136-1152.

Fraser G. Pityrosporum pachydermatis Weidman of canine origin. Trans Br My-col Soc 1961;44:441–448.

Freel K. C., Friedrich A., Schacherer J. Mitochondrial genome evolution in ye-asts: an all-encompassing view. FEMS Yeast Res. [internet] 2015 [cited Jul 15];15 (4):fov023. Available from: https://pubmed.ncbi.nlm.nih.gov/25969454/. doi:10.1093/femsyr/ fov023.

Gandra RF, Gambale W, de Cássia Garcia Simão R et al. *Malassezia* spp. in acoustic meatus of bats (Molossus molossus) of the Amazon Region, Brazil. Myco-pathol 2008;165:21–26.

Glatz, M., Buchner, M., Von Bartenwerffer, W., Schmid-Grendelmeier, P., Worm, M., Hedderich, J. et al. *Malassezia* spp.-specific immunoglobulin E level is a marker for severity of atopic dermatitis in adults. Acta Derm. Venereol. [internet] 2015 [cited Jul 15];95(2):191-196. Available from: https://medicaljournalssweden.se/actadv/article/view/5645 doi: 10.2340/00015555-1864.

Gordon, M. A. Lipophilic yeastlike organisms associated with tinea versicolor. J. Invest. Dermatol. 1951;17:267-272. doi: 10.1038/jid.1951.93

Gougerot H, Carteaud A. Papillomatose pigmentée innominée. Cas pour diagnos-tic. Bull Soc Fr Dermatol Syphil 1927;34:712-719.

Grahl N., Dinamarco T. M., Willger S. D., Goldman G. H., Cramer R. A. Asper-gillus fumigatus mitochondrial electron transport chain mediates oxidative stress ho-meostasis, hypoxia responses and fungal pathogenesis: mitochondrial respiration andAspergillus pathogenesis. Mol. Microbiol. [internet] 2012 [cited Jul 15];84:383-399. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3323727/pdf/nihms362511.pdf. doi:10.1111/j.1365- 2958.2012.08034.

Guého E, Midgley G, Guillot J. The genus Malassezia with description of fournew species. Antonie van Leeuwenhoek 1996;69:337-355.

Guillot J, Guého E, Lesourd M et al. Identification of Malassezia species. A prac-tical approach. J Mycol Méd 1996;6:103–110.

Gupta AK, Kohli Y, Faergemann J et al. Epidemiology of Malassezia yeasts asso-ciated with pityriasis versicolor in Ontario, Canada. Med Mycol 2001;39:199–206.

Gupta, A. K., Bluhm, R., Summerbell, R. Pityriasis versicolor. J. Eur.Acad. Dermatol. Venereol. [internet] 2002[cited Jul 15];16:19-33. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1046/j.1468-3083.2002.00378.x. doi: 10.1046/j.1468-3083.2002.00378.

Gustafson BA. Otitis externa in the dog. A bacteriological and experimentalstudy. Thesis. Royal Veterinary College Sweden, Stockholm; 1955.

Hafez M., Hausner G. Homing endonucleases: DNA scissors on a mission. Ge-nome 2012;55: 553-569.

Hao, W. From genome variation to molecular mechanisms: What we have learned fromyeast mitochondrial genomes?. Front. Microbiol. 2022;13:806575.

Hausner G. Fungal mitochondrial genomes, plasmids and introns. In: Arora D. K., Khachatourians G. G. editores. Applied Mycology and Biotechnology. Amsterdam:Elsevier; 2003. p.101–131.

Hausner G., Hafez M., Edgell D. R. Bacterial group I introns: mobile RNA ca-talysts. Mobile DNA 2014;5:8.

Hirai A, Kano R, Makimura K, Duarte ER, Hamdan JS, Lachance MA, Yamaguchi H, Hasegawa A. *Malassezia* nana sp. nov., a novel lipid-dependent yeast species isola-ted from animals. Int J Syst Evol Microbiol. [internet] 2004 Mar[cited Jul 15];;54(Pt 2):623-627. doi: 10.1099/ijs.0.02776-0. Pudmed PMID: 15023986.

Hobi S, Cafarchia C, Romano V, Barrs VR. *Malassezia*: Zoonotic Implications, Paral-lels and Differences in Colonization and Disease in Humans and Animals. J Fungi (Basel). [internet] 2022 Jul 4[cited Jul 15];8(7):708. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9324274/pdf/jof-08-00708.pdf. doi: 10.3390/jof8070708. Pudmed PMID: 35887463; Pubmed Central PMCID:PMC9324274.

Honnavar, P., Prasad, G. S., Ghosh, A., Dogra, S., Handa, S., Rudramurthy, S. M. *Malassezia* arunalokei sp. nov., a Novel Yeast Species Isolated from Seborr-heic Dermatitis Patients and Healthy Individuals from India. J. Clin. Microbiol. [internet] 2016 [cited Jul 15];54:1826–1834. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4922115/pdf/zjm1826.pdf. doi: 10.1128/JCM.00683-16

Huszcz G.B; Teixeira G.E; Rodrigues-Oliveira I.H; Rocha R.R; Oliveira RC; Takao PS et al. Five mitochon- drial genomes of the lipophilic fungus *Malassezia* (*Malassezia*les, Exobasidiomyce- tes). RevistaFT, 2023;27(124):1-2.

Hyde, K. D., Xu, J., Rapior, S., Jeewon, R., Lumyong, S., Niego, A. G. T. et al. The amazing potential of fungi: 50 ways we can exploit fungi industrially. FungalDivers. [internet] 2019 [cited Jul 15];97:1-136. Available from: https://link.springer.com/article/10.1007/s13225-019-00430-9. doi: 10.1007/s13225-019-

00430-9.

Ianiri G, Heitman J. Approaches for Genetic Discoveries in the Skin Commensal and Pathogenic *Malassezia* Yeasts. Front Cell Infect Microbiol. [internet] 2020 Aug 7[cited Jul 15];10:393. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7426719/pdf/fcimb-10-00393.pdf. doi: 10.3389/fcimb.2020.00393. Pubmed PMID: 32850491; Pubmed Central PMCID: PMC7426719.

Ingham E, Cunningham AC. Malassezia furfur. J Med Vet Mycol [internet] 1993 [cited Jul 15];;31:265-288.

Islamoglu, Z. G. K. Second-to-fourth digit ratio and seborrheic dermatitis in males: a cross-sectional study. Bras. Dermatol. [internet] 2019 [cited Jul 15];94:327-330. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6668940/pdf/abd-94-03-0327.pdf. doi: 10.1590/abd1806-4841.20198578.

James T. Y., Pelin A., Bonen L., Ahrendt S., Sain D., Corradi N. et al. Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia.Curr. Biol. [internet] 2013[cited Jul 15];23:1548-1553. Available from: https://www.sciencedirect.com/science/article/pii/S0960982213007823?via%3Dihub doi: 10.1016/j.cub.2013.06.057.

James T. Y., Stajich J. E., Hittinger C. T., Rokas A. Toward a fully resolved fungal tree of life. Annu. Rev. Microbiol. [internet] 2020 [cited Jul 15];74:291-313. Available from: https://www.annualreviews.org/doi/10.1146/annurev-micro-022020-051835?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed. doi:10.1146/annurev-micro-022020-051835.

Jansen, R. K. et al. Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol. 2005;395:348-384.

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder:fast model selection for accurate phylogenetic estimates. Nature Methods 2017;4(6):587-589.

Kaniak-Golik A., Skoneczna A. Mitochondria–nucleus network for genome sta-bility. Free Radical Biol. Med. 2015;82:73-104.

Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013;30(4):772–780.

Kim, G. K. Seborrheic Dermatitis and *Malassezia* species: How Are They Rela-ted? J. Clin. Aesthet. Dermatol. 2009;2:14-17.

Koh LF, Ong RY, Common JE. Skin microbiome of atopic dermatitis. Allergol Int. [internet] 2022 Jan [cited Jul 15];71(1):31-39. Available from: https://www.sciencedirect.com/science/article/pii/S1323893021001404?via%3Dihub. doi: 10.1016/j.alit.2021.11.001. Epub 2021 Nov 24. Pubmed PMID: 34838450.

Korovesi, A. G., Ntertilis, M. & Kouvelis, V. N. Mt-rps3 is an ancient gene which pro- vides insight into the evolution of fungal mitochondrial genomes. Mol. Phylogenet.Evol. 2018;127: 74-86.

Kulik T., Ostrowska A., Buśko M., Pasquali M., Beyer M., Stenglein S. et al. Development of an FgMito assay: a highly sensitive mitochondrial based qPCR as- say for quantification of Fusarium graminearum sensu stricto. Int. J. Food Microbi-ol. [internet] 2015[cited Jul 15]; 210:16-23. Available from: https://www.sciencedirect.com/science/article/pii/S0168160515300362?via%3Dihub doi: 10.1016/j.ijfoodmicro.2015.06.012.

Kulik, T., Bilska, K., Zelechowski, M. Promising perspectives for detection,identification, and quantification of plant pathogenic fungi and oomycetes through targeting mitochondrial DNA. Int. J. Mol. Sci. [internet] 2020 [cited Jul 15];21:2645. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7177237/pdf/ijms-21-02645.pdf. doi: 10.3390/ijms21072645.

Kulik T., Van Diepeningen A. D., Hausner G. Editorial: the significance of mi- togenomics in mycology. Front. Microbiol. 2021;11:628579.

Kurniadi I, Hendra Wijaya W, Timotius KH. *Malassezia* virulence factors and their rolein dermatological disorders. Acta Dermatovenerol Alp Pannonica Adriat. 2022 Jun;31(2):65-70. PMID: 35751534.

Lai X, Cao L, Tan H et al.Fungal communities from methane hydrate-bearingdeep-sea marine sediments in South China Sea. ISME J 2007;1:756–762.

Lang B. F., Laforest M.-J., Burger G. Mitochondrial introns: a critical view. Trends Genet. [internet] 2007[cited Jul 15];23:119-125. Available from: https://www.sciencedirect.com/science/article/pii/S0168952507000248?via%3Dihub. 10.1016/j.tig.2007.01.006.

Li, Q., Ren, Y., Shi, X., Peng, L., Zhao, J., Song, Y. et al. Comparative mito- chondrial genome analysis of two ectomycorrhizal fungi (Rhizopogon) reveals dyna-mic changes of intron and phylogenetic relationships of the subphylum agaricomyco-tina. Int. J. Mol. Sci. [internet] 2019[cited Jul 15];20:5167. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6829451/pdf/ijms-20-05167.pdf. doi: 10.3390/ijms20205167.

Li Y., Steenwyk J. L., Chang Y., Wang Y., James T. Y., Stajich J. E. et al. Agenome-scale phylogeny of the kingdom Fungi. Curr. Biol. 2021;31:1653.e5-1665.e5.

Limon JJ, Tang J, Li D *et al. Malassezia* is associated with Crohn's disease and exacer-bates colitis in mouse models. Cell Host Microbe. 2019;25:377–88.

Liu W., Cai Y., Zhang Q., Shu F., Chen L., Ma X. et al. Subchromosome-Scalenuclear and complete mitochondrial genome characteristics of morchella crassipes. Int. J. Mol. Sci. 2020;21:483.

Malassez L. Note sur le champignon du pityriasis simple. Arch Physiol 1874;1:451-459.

Malina, C., Larsson, C., and Nielsen, J. Yeast mitochondria: an overview of mi-tochondrial biology and the potential of mitochondrial systems biology. FEMS Ye- ast Res. 2018;18:40.

Marcon MJ, Powell DA. Human infections due to *Malassezia* spp. Clin Microbi-ol Rev 1992;2:101–119.

Medina R, Franco MEE, Bartel LC, Martinez Alcántara V, Saparrat MCN, Balatti PA.Fungal Mitogenomes: Relevant Features to Planning Plant Disease Management. Front Microbiol. [internet] 2020 May 29[cited Jul 15];11:978. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7272585/pdf/fmicb-11-00978.pdf doi: 10.3389/fmicb.2020.00978. Pubmed PMID:32547508; Pubmed Central PMCID: PMC7272585.

Mendoza H., Perlin M. H., Schirawski J. Mitochondrial inheritance in phytopa-thogenic fungi—everything is known, or is it? Int. J. Mol. Sci. 2020;21:3883.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 2020;37(5):1530-1534.

Moreno-Indias, I. et al. Statistical and machine learning techniques in human microbio-me studies: Contemporary challenges and solutions. Front. Microbiol. 2021;12:635781.

Mudadu MA, Zerlotini A. Machado: Open source genomics data integration framework. Gigascience. [internet] 2020 Sep 14[cited Jul 15];9(9):giaa097. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7490629/pdf/giaa097.pdf. doi: 10.1093/gigascience/giaa097. Pubmed PMID:32930331; Pubmed PMCID: PMC7490629.

Mungall C.J., Emmert D.B., FlyBase Consortium. A Chado case study: an onto- logy-based modular schema for representing genome-associated biological informati-on. Bioinformatics 2007;23:i337–i346.

Muñoz-Gómez S. A., Wideman J. G., Roger A. J., Slamovits C. H. The origin of mitochondrial cristae from alphaproteobacteria. Mol. Biol. Evol. 2017;34:943-956.

Nakabayashi, A., Sei, Y., and Guillot, J. Identification of *Malassezia* species iso-lated from patients with seborrhoeic dermatitis, atopic dermatitis, pityriasis versico- lor and normal subjects. Med. Mycol. 2000;38:337–341.

Narang T, Dogra S, Kaur I, Kanwar A J et al. Malassezia and psoriasis: Koebner's phenome-non or direct causation? J Eur Ac Dermatol Venereol 2007;21:1111–1112.

Naranjo-Ortiz, M. A., and Gabaldón, T. Fungal evolution: cellular, genomic andmetabolic complexity. Biol. Rev. 2020;95:brv.12605.

Oh, J., Byrd, A. L., Deming, C., Conlan, S., Program, N. C. S., Kong, H. H. et al. Biogeography and individuality shape function in the human skin metageno-me. Nature 2014;514:59–64.

Oh, J., Byrd, A. L., Park, M., Program, N. C. S., Kong, H. H., Segre, J. A. Temporal Stability of the Human Skin Microbiome. Cell [internet] 2016[cited Jul 15];165:854–866. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4860256/pdf/nihms776449.pdf. doi:10.1016/j.cell.2016.04.008

Ota M, Huang PT .Sur les champignons du genre Pityrosporum Sabouraud. AnnParasitol Hum Comp 1933;11:49–69.

Palmer, J. D. Contrasting modes and tempos of genome evolution in land plant organel-les. Trends Genet. 1990;6:115–120.

Panja G. The Malassezia of the skin, their cultivation, morphology and species. Trans 7th Cong Far Eastern Ass Trop Med 1927;2:442–455.

Pantou, M. P., Kouvelis, V. N. & Typas, M. A. The complete mitochondrial genome of Fusarium oxysporum: Insights into fungal mitochondrial evolution. Gene 2008;419:7–15.

Park M, Park S, Jung WH. Skin Commensal Fungus *Malassezia* and Its Lipases. J Mi- crobiol Biotechnol. [internet] 2021 May 28[cited Jul 15];31(5):637-644. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9705927/pdf/jmb-31-5-637.pdf doi: 10.4014/jmb.2012.12048. Pubmed PMID: 33526754; Pubmed Central PMCID: PMC9705927.

Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F et al. Sequenceserver: A Modern Graphical User Interface for Cus- tom BLAST Databases. Mol Biol Evol. [internet] 2019 Dec 1[cited Jul 15];36(12):2922-2924. Available from: https://academic.oup.com/mbe/article/36/12/2922/5549819. doi: 10.1093/molbev/msz185. Pubmed PMID: 31411700; Pubmed Central PMCID: PMC6878946.

Redline RW, Dahms BB.Malassezia pulmonary vasculitis in an infant on long-term intralipid therapy. N Engl J Med 1981;305:1395–1399.

Renker C, Alphei J, Buscot F. Soil nematodes associated with the mammal patho-genic fungal genus *Malassezia* (Basidiomycota: Ustilaginomycetes) in Central Euro- pean forests. Biol Fertil Soils 2003;37:70–72.

Rhimi W, Theelen B, Boekhout T, Otranto D, Cafarchia C. *Malassezia* spp. Yeasts of Emerging Concern in Fungemia. Front Cell Infect Microbiol. [internet] 2020 Jul 28[cited Jul 15];10:370. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7399178/pdf/fcimb-10-00370.pdf. doi: 10.3389/fcimb.2020.00370. Pubmed PMID: 32850475; Pubmed Central PMCID: PMC7399178.

Ro, B. I., and Dawson, T. L. The role of sebaceous gland activity and scalp mi- crofloral metabolism in the etiology of seborrheic dermatitis and dandruff. J. Invest. Dermatol. Symp. Proc. 2005;10:194–197. doi: 10.1111/j.1087-0024.2005.10104.x

Robin C. Histoire naturelle des végétaux parasites qui croissent sur l'homme etles ani-maux

vivants. France: Hachette Livre BNF; 1853.

Baillière, Paris Rosenberg EW, Belew PW. Improvement of psoriasis of the scalp with ketocona-zole. Arch Dermatol 1982;118:370–371.

Rubinoff D., Holland B. S. Between two extremes: mitochondrial DNA is nei-ther the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 2005;54:952-961.

Sabouraud R. Maladies du cuir chevelu. II. Les maladies desquamatives. Français: Hachette Livre BNF;1904.

Masson& Cie, Paris Saunte DML, Gaitanis G, Hay RJ. *Malassezia*-Associated Skin Diseases, the Use of Di-agnostics and Treatment. Front Cell Infect Microbiol. [internet] 2020 Mar 20[cited Jul 15];10:112. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7098993/pdf/fcimb-10-00112.pdf. doi: 10.3389/fcimb.2020.00112. Pubmed PMID: 32266163; Pubmed Central PMCID: PMC7098993.

Schmidt-Dannert C. Biocatalytic portfolio of Basidiomycota. Curr. Opin. Chem. Biol. 2016;31:40-49.

Silva, V., Fischman, O., and De Camargo, Z. P. Humoral immune response to *Malassezia* furfur in patients with pityriasis versicolor and seborrheic dermatitis. Mycopathologia 1997;139:79-85.

Simmons RB, Guého E. A new species of Malassezia. Mycol Res 1990;94:1146–1149.

Sizun J, Karangwa A, Giroux JD et al. Malassezia furfur-related colonization andinfection of central venous catheters. A prospective study in a pediatric intensive careunit. Int Care Med 1994;20:496–499.

Slooff WC. Genus 6 Pityrosporum Sabouraud. In: Lodder J, editor. The Yeasts, ataxonomic study. 2rd. ed. North-Holland, Amsterdam; 1970. p 1167–1186.

Smith D. R. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? Briefings Funct. Genomics 2015;15:47–54.

Spatz M, Richard ML. Overview of the potential role of *Malassezia* in gut health and di-sease. Front Cell Infect Microbiol. 2020;10:201.

Sugita, T., Tajima, M., Takashima, M., Amaya, M., Saito, M., Tsuboi, R. et al. A new yeast, *Malassezia* yamatoensis, isolated from a patient with seborrheic derma-titis, and its distribution in patients and healthy subjects. Microbiol. Immunol. [internet] 2004[cited Jul 15];48:579-583. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1348-0421.2004.tb03554.x.
doi: 10.1111/j.1348-0421.2004.tb03554.x

Walker G. M., White N. A. Introduction to fungal physiology. In: Fungi, Kavanagh K., editor. Chichester: John Wiley & Sons, Ltd; 2005. p. 1–34.

Theelen, B., Cafarchia, C., Gaitanis, G., Bassukas, I. D., Boekhout, T., and Dawson, T. L.Jr. *Malassezia* ecology, pathophysiology, and treatment. Med. Mycol. 2018;56:S10–S25.

Theelen, B., Cafarchia, C., Gaitanis, G., Bassukas, I. D., Boekhout, T., and Dawson, T. L. Corrigendum: *Malassezia* ecology, pathophysiology, and treatment. Med.Mycol. 2019;57:e2.

Unna PG. Die Färbung der Mikroorganismen im Horngewebe. Monatschr FPrakt Dermatol 1891;13:225-237.

Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 2011;27(2):171-180.

Valach, M., Farkas, Z., Fricova, D., Kovac, J., Brejova, B., Vinar, T. et al. Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. Nu- cleic Acids Res. [internet] 2011[cited Jul 15];39:4202–4219. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3105423/pdf/gkq1345.pdf. doi: 10.1093/nar/gkq1345

Verma S., Shakya V. P. S., Idnurm A. (2018). Exploring and exploiting the connection between mitochondria and the virulence of human pathogenic fungi. Virulence 2018;9:426-446.

Vijaya Chandra SH, Srinivas R, Dawson TL Jr, Common JE. Cutaneous *Malassezia*: Commensal, Pathogen, or Protector? Front Cell Infect Microbiol. [internet] 2021 Jan 26 [cited Jul 15];10:614446.
Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7870721/pdf/fcimb-10-614446.pdf.
doi: 10.3389/fcimb.2020.614446. Pubmed PMID: 33575223; Pubmed Central PMCID: PMC7870721.

Wananukul, S., Chindamporn, A., Yumyourn, P., Payungporn, S., Samathi, C., Poo-vorawan, Y. *Malassezia* furfur in infantile seborrheic dermatitis. Asian Pac. J. Allergy Immunol. 2005;23:101–105.

Wang QM, Theelen B, Groenewald M, Bai FY, Boekhout T. Moniliellomycetes and Malasseziomycetes, two new classes in Ustilaginomycotina. Persoonia-MolecularPhylogeny and Evolution of Fungi. 2014;33(1):41-47.

Wang, G., Lin, J., Shi, Y., Chang, X., Wang, Y., Guo, L. et al. Mitochondrial genome in Hypsizygus marmoreus and its evolution in Dikarya. BMC Genomics [internet] 2019 [cited Jul 15];20:765. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6805638/pdf/12864_2019_Article_6133.pdf. doi: 10.1186/s12864-019-6133-z.

Weidman FD. Exfoliative dermatitis in the Indian rhinoceros (Rhinoceros unicor-nis), with description of a new species: Pityrosporum pachydermatis. In: Fox H, editor. Rep Lab Museum Comp Pathol Zoo Soc Philadelphia Philadelphia: University of California; 2005. p. 36–43.

Wikramanayake TC, Borda LJ, Miteva M, Paus R. Seborrheic dermatitis-Lookingbeyond *Malassezia*. Exp Dermatol. [internet] 2019 Sep [cited Jul 15];28(9):991-1001. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1111/exd.14006.
doi: 10.1111/ exd.14006.

Williamson D. The curious history of yeast mitochondrial DNA. Nat. Rev. Ge-net. 2002;3:475-481.

Wu, G., Zhao, H., Li, C., Rajapakse, M. P., Wong, W. C. et al. Genus-wide comparative genomics of *Malassezia* delineates its phylogeny, physiology, and niche adaptation on human skin. PLoS Genetics 2015;11:e1005614.

Wu, B., Hussain, M., Zhang, W., Stadler, M., Liu, X., and Xiang, M. Current in-sights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. Mycology [internet] 2019 [cited Jul 15];10:127–140. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691916/pdf/TMYC_10_1614106.pdf.
doi: 10.1080/21501203.2019.1614106

Xu H, Li H. Acne, the Skin Microbiome, and Antibiotic Treatment. Am J Clin Derma-tol. [internet] 2019 Jun [cited Jul 15];20(3):335-344. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6534434/pdf/nihms-1518421.pdf.
doi: 10.1007/s40257-018-00417-3. Pubmed PMID: 30632097; Pubmed Central PMCID: PMC6534434.

Zardoya R. Recent advances in understanding mitochondrial genome diversity. F1000Res 2020;9:270.

Zhang Y, Yang G, Fang M, Deng C, Zhang KQ, Yu Z, Xu J. Comparative Analyses of Mitochondrial Genomes Provide Evolutionary Insights Into Nematode-Trapping Fungi. Front Microbiol. [internet] 2020 Apr 15[cited Jul 15];11:617. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7174627/pdf/fmicb-11-00617.pdf
doi: 10.3389/fmicb.2020.00617. Pubmed PMID: 32351475; Pubmed Central PMCID: PMC7174627.

Zhao, X. Q. et al. Complete mitochondrial genome of the aluminum-tolerant fungus Rhodotorula taiwanensis RS1 and comparative analysis of Basidiomycota mitochon-drial genomes. Microbiologyopen 2013;2:308–317.

ANEXOS

ISSN 1678-0817 Qualis B2
Revista Cientí ca de Alto Impacto.

FIVE MITOCHONDRIAL GENOMES OF THE LIPOPHILICFUNGUS MALASSEZIA (MALA SSEZIALES EXOBASIDIOMYCETES)

Ciências Biológicas, Edição 124 JUL/23 SUMÁRIO / 12/07/2023

REGISTRO DOI: 10.5281/zenodo.8140802

Gabriel Barbosa Huszcz[1,2], Gledison Eric Teixeira [1], Igor Henrique Rodrigues-Oliveira[3], Renan Rodrigues Rocha[2,3], Rafaela de Campos Oliveira[2], Paulo Sallarola Takao[2], David Aciole Barbosa[1], Daniela Leite Jabes[1],Karine Frehner Kavalco[3], Rubens Pasa[3], Fabiano Bezerra Menegidio[1,2]

Abstract: Fungi of the *Malassezia* genus are the most prevalent eukaryotic organisms in the natural skin mycobiome of animals and humans. The genus is found in various parts of the body, causing common skin diseases such as pityriasis versicolor and seborrheic dermatitis, and its high adaptability to the human body and prevalence among other eukaryotes makes it unique. However, even with eighteen known species and genomes available, studies on their phylogeny and basic genomic information are still scarce. To this day, only nine species have their mitochondrial genome described and assembled. In this study, we describe for the mtDNA of *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina*, *Malassezia caprae* and six new mitogenomes of *Malassezia pachydermatis*. In addition, to better understand thephylogenetic relationships, we mined 35 mitogenomes belonging to 9 species of *Malassezia*, along with the mitogenomes gathered in this study, and reconstructed the phylogeny based on mtDNA. The results of this study providegenomic variation information and enhance the understanding of the *Malassezia* genus. They could later deliver highly valuable new insight into data for phylogenetic analysis and population genetics.

Keywords: Mitochondrial genome, *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina*, Malassezia caprae, *Malasseziapachydermatis*, Phylogeny.

Introduction

Yeasts of the *Malassezia* genus are present in the microbiota of many animals, including humans, living as commensal organisms on the skin and scalp (Findley et al., 2020). However, in some cases, they can

become pathogenic under the
in uence of predisposing factors that involve both the conditions of the skin microenvironment and changes in the host's immune system (Theelen et al., 2018; Aykut et al., 2019; Limon et al., 2019; Saunte et al., 2020; Spatz and Richard 2020). The genus groups basidiomycetic, lipophilic (*M. pachydermatis*) and/or lipodependent yeasts that show asexual reproduction by unipolar or sympodial budding (*M. sympodialis*) from the parent cell.

Another important feature is the possibility of forming pseudohyphae.

Differentiation of species in the traditional way is done through lipid and fatty acid assimilation tests (biochemical methods), also considering the characteristics of micromorphology. Currently, molecular biology methodologies have been considered of great importance for the taxonomic grouping of the *Malassezia* genus. The support of molecular tools is very important, since the morphological characteristics between species are very subtle, which can generate doubts and errors. Study of the D1/D2 regions of the 28S ribosomal RNA gene (rRNA28S) and the ITS (Internal Transcribed Spacer) regions have helped in the differentiation of *Malassezia* species (Batra et al., 2005). In 1996, based on studies based on morphology, physiology and molecular biology, there was a change in the classi cation of the genus *Malassezia* in seven species. As of 2002, another seven new species of *Malassezia* were cited in the literature. Currently, 18 species are recognized for the genus, with only 17 having available genomes (*M. arunalokei*, *M. caprae*, *M. cuniculi*, *M. dermatis*, *M. equina*, *M. furfur*, *M. globosa*, *M.japonica*, *M. nana*, *M. obtusa*, *M. pachydermatis*, *M. restricta*, *M. sloof ae*, *Malassezia sp.*, *M. sympodialis*, *M. vespertilionis* and *M. yamatoensis*) and 9 having mitochondrial genomes (*M. furfur*, *M. globosa*, *M. japonica*, *M. obtusa*,

M. pachydermatis, *M. restricta*, *M. sloof ae*, *M. sympodialis* and *M. yamatoensis*).

In this study, we describe the mitochondrial genome of ve species of Malasseziales: *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina* and *Malassezia caprae*, also six new mitogenomes of *Malassezia pachydermatis*. In addition, to better understand the phylogenetic relationships, we mined 35 mitogenomes belonging to 9 species of *Malassezia*, along with the mitogenomes gathered in this study, and reconstructed the phylogeny based on mtDNA.

Materials & Methods

Mitogenome assembly and annotation

We obtained Whole Genome Sequencing (WGS) raw data of *Malassezia vespertilionis* (SRR6206152; Lorch et al., 2018), *Malassezia dermatis* (DRR043255; Sugita et al., 2002), *Malassezia nana* (DRR043258; Hirai et al., 2004), *Malassezia equina* (SRR2136627; Wu et al.,2015) and *Malassezia caprae* (SRR2132347; Wu et al., 2015) from the Sequence Reads Archive (SRA) NCBI.

The reads were submitted to a work ow previously used and validated in sh mitogenome assemblies by Rocha-Reis et al. (2020), Resende et al. (2020) and Pasa et al. (2021), with minor modi cations to adapt to the assembly of fungal mitogenomes. We imported the raw data into the Galaxy Europe platform (https://usegalaxy.eu/) and used NOVOplasty v4.3.1 (Dierckxsens et al.,2017) to assemble the mitochondrial genomes. As a seed, we used the complete mitogenome of M. japonica (KY911090.1). We annotated the

sequences obtained on Mitos2 (http://mitos2.bioinf.uni-leipzig.de/; Arab et al., 2017) using Gene Code 3
– Yeast and RefSeq 89 Fungi.
To validate our work ow, we also performed the assembly and annotation of six new mitogenomes from *Malassezia pachydermatis* WGS libraries (SRR12046913, SRR12005311, SRR12005312, SRR12005313, SRR12005314 and SRR2135029). In addition, the 34 mitogenomes of the 9 species of the *Malassezia* genus available at the NCBI (*M. furfur*, *M. globosa*, *M. japonica*, *M. obtusa*, *M. pachydermatis*, *M. restricta*, *M. sloof ae*, *M. sympodialis* and *M. yamatoensis*) were annotated according to the procedure described above.

The circular map of the new mitogenomes was drawn using the online softwareGenomeVx (Conant and Wolfe, 2008). We used the Blast Ring Image Generator (BRIG) (Alikhan et al., 2011) to perform a comparative Blast analysis of available representative Malassezia mitogenomes against our mitogenome pools.

Mitochondrial phylogenetic analysis

To perform the phylogenetic analyses, we manually extracted the sequences of the 15 proteincoding genes (PCGs) of the 11 mitogenomes assembled in this work and added to the dataset the PCGs of the 34 mitogenomes of other Malassezia species available at the NCBI, addition to the *Rhodotorula mucilaginous* (MF694646.1) mitogenome as an outgroup. We aligned the sequences with MAFFT v7 (Katoh & Standley, 2013), available on the Galaxy Europe online platform (Afgan et al., 2018), and concatenated the alignments in SequenceMatrix v1.8 (Vaidya et al., 2011). We constructed the phylogeny by the Maximum Likelihood (ML) method in the IqTree

v2.1.2 software (Nguyen et al., 2015) using 1000 ultra-fast bootstrap repeats and the evolutionary model GTR+F+I+G4 estimated by the IqTree ModelFinder (Kalyaanamoorthy et al., 2017).

Data Description

Characterization of mitochondrial genomes of Malassezia species

The mitogenome of the ve Malassezia species are circular DNA molecules with a total length ranging from 27,284 bp to 40,970 bp. Mitogenome sizes varied widely among the ve newly assembled mitochondrial genomes, with *M. vespertilionis* having the largest among the new mtDNA and *M. caprae* being the smallest mitogenome ever described for the genus. Furthermore, *M. equina* becomes the second smallest mitogenome described for the genus, followed bythe already described *M. japonica*. However, *M. dermatis* and *M. nana* presented a mitogenome with a size close to the average of other species of the genus. Likewise, the new *M. pachydermatis* mitogenomes were between 35,817 bp and 35,829 bp in size, being slightly larger than the *M. pachydermatis* strain CBS1879 (KY911092.1) mitogenome previously described. A comparison between the size of the new mitogenomes described in this work and those available in the NCBI database can be seen in Supplementary Tables 1.

The GC concentration of the mitogenome differs among species, which may beaffected by mutation bias, selection, and biases of reconstitution-related DNA repair (Li et al., 2018). The low G + C content observed is similar to other fungal mitogenomes in Malassezia (Supplementary Table 1). The GC content in the ve

mitogenomes ranged from 23.4% to 33.7%, among which the GC content in thegenome of *M. caprae* was the highest.

The base composition of the mitochondrial genome of *M. caprae* was estimated as 33.8% A, 17.2%C, 16.5% G, and 32.5% T, thus presenting a bias towards AT (66.3%) and the GC content was 33.7%. *M. equina* presented the composition of 33.4% A, 33.7% T, 16.0% C and 16.1% G, with a bias towards AT (67.1%) and GC content of 32.9%. *M. nana* presented the composition of 33.6% A, 34.5% T, 15.8% C and 16.1% G, with a bias towards AT (68.1%) and GC content of 31.9%.*M. vespertilionis* presented the composition of 36.8% A, 39.8% T, 11.5% C and 11.8% G, with a bias towards AT (76.6%) and GC content of 23.4%. Finally, *M. dermatis* presented a composition of 34.1% A, 34.3% T, 15.6% C and 16.0% G, with a bias towards AT (68.4%) and GC content of 31.6%.

In general, mitogenomic coding is associated with the mitochondrial translation apparatus, electron transport, and oxidative phosphorylation. The potential coding genes (PCGs) of Malassezia species included 14 PCGs related to oxidative phosphorylation [NAD dehydrogenase (nad1-6 and nad4L), cytochrome oxidases (cox1, cox2 and cox3), cytochrome b (cob), and 3 ATP synthases (atp6, atp8 and atp9), ribosomal protein S3 (rps3)] and 2 ribosomal RNA genes (rns and rnl). Only *M. equina*, *M. caprae* and *M. vespertilionis* presented 14 PCGs, due to the absence of the atp9 gene. This distribution is consistent with that observed in typical fungal mitogenomes, such as those belonging to *Ascomycota* and *Basidiomycota* (*Dikarya* Subkingdom). In total, the coding regions containing the15 PCGs described occupied between 13,725 bp and 14,103 bp in the two mitogenomes presented (*M. nana* and *M. dermatis*, respectively). Likewise, the 15 PCGs occupied from

14,037 bp to 14,208 bp among the 6 new mitogenomes of *M. pachydermatis* described here, in addition to 12,735 bp and 14,784 bp among the mitogenomes available in the NCBI database. In mitogenomes that presented 14 PCGs, the coding regions occupied between 13,437 bp and 14,079 bp (*M. equina* and *M. caprae*, respectively). Other existing regions in mitogenomes are tRNAs, rRNA genes and intergenic regions. More details on these regions contained in mitogenomes are available in Supplementary Table 1.

Except for *M. vespertilionis*, all new mitogenomes showed nad4 and atp6 in the light chain. Likewise, a duplication of the atp9 gene in *M. dermatis*, *M. nana* and in the new *M. pachydermatis* mitogenomes was also observed in the light chain. In *M. vespertilionis*, only the cob gene was observed in the light chain. Apart from those already described, all other genes were observed in the heavy chain for all new mitogenomes. More details on the direction of PCGs can be found in Supplementary Tables 1-2.

Besides *M. caprae* and *M. equina*, where 18 tRNA genes were identi ed in the mitogenomes, the other species characterized in this work presented between 20 and 25 tRNAs, encoding the 20 standard amino acids. The length of tRNA genes in the ve mitogenomes ranged from 71 bp to 86 bp. Each of the ve *Malassezia* mitogenomes involved 2 rRNA genes: the small subunit ribosomal RNA gene (rns) and the large subunit ribosomal RNA gene (rnl). No signicant difference was identied in the number of rRNA genes in the vespecies, and the total length of the genes.

Phylogenetic Analysis

We obtained a robust evolutionary tree topology (Figure 1) with all recovered clades well supported. Our phylogeny is mostly accordant

with a prior study based on sequences of six mitochondrial and nuclear genes (Wang et al., 2014). The tree has shown two main clades comprising a distinct group of species. As in the literature, our data reached a clade encompassing *M. caprae*, *M. dermatis*, *M. sympodialis*, *M. equina*, and *M. nana*. This clade is near related to *M. pachydermatis* clade. Another well-marked and related clade comprises *M. globosa* and *M. restricta* lineages.

The other main clade assembles *M. slof ae*, *M. vespertilionis*, *M. japonica*, *M. yamatoensis*, *M. obtusa*, and a species complex formed by samples identied as *M. furfur*. Our analysis shows two clades of *M. furfur* with 100% bootstrap condence and a single strain diverging early. The bootstrap condence of this strain with the two largest clades reached 100% of condence.

The main difference between our analysis and Wang et al. (2014) study was the position of *M. restricta*, *M. globosa*, and *M. pachydermatis*. In our three (Figure 1), these species are closest to the rst mentioned clade, though previous analyzes grouped them with the second clade (Wang et al., 2014). In the phylogenetic analysis, the novel *Malassezia pachydermatis* mitogenome had clustered consistently with *M. pachydermatis* strain CBS1879 (KY911092.1). The same had happened with other species with multiple mitogenomes available.

Conclusion

The methodology used to assembly the mitochondrial genomes of *Malassezia vespertilionis*, *Malassezia dermatis*, *Malassezia nana*, *Malassezia equina* and *Malassezia caprae* proved to be satisfactory and allowed the expanded the phylogeny of *Malasseziales* through mtDNA. We were also able to validate our work ow by assembling 6

new *Malassezia pachydermatis* mitogenomes that demonstrated high agreement with the mitogenome already available. Furthermore, the study of the complete mitochondrial genome proves to be atool with potential to solve taxonomic problems and help in understanding theevolutionary and ecological relationships of this group.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

RCO, IHRO, GH, GET, PST, RRR, DAB, DLJ, KFK, RP and FBM contributed to the conceptualization, investigation and design of the study. RCO, IHRO, GH, GET, PST, RRR, DAB, DLJ, KFK, RP and FBM performed the bioinformatics/statistics analysis and were responsible for data validation. FBM and DLJ provided resources to carry out the work. FBM, DLJ, KFK and RP were responsible for the administration and supervision of the research project. FBM, DAB, RCO, IHRO, PST and RRR wrote the rst draft of the manuscript. FBM, DLJ, KFK and RP wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Acknowledgments

This study was nanced in part by scholarship grants from Coordenação deAperfeiçoamento de Pessoal de Nível Superior – Brasil – CAPES (www.capes.gov.br/) and Conselho Nacional de Desenvolvimento Cientí co e Tecnológico – CNPq

(www.gov.br/cnpq/pt-br).

Supplementary Material

The Supplementary Material for this article can be found online at: https://doi.org/10.6084/m9.figshare.23666289.v1

Reference

Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses:2018 update. Nucleic acids research. 2018;46(W1):W537-W544.

Al Arab M, Zu Siederdissen CH, Tout K, Sahyoun AH, Stadler PF, Bernt M. Accurate annotation of protein-coding genes in mitochondrial genomes.Molecular phylogenetics and evolution. 2017;106: 209-216.

Alikhan, NF., Petty, N.K., Ben Zakour, N.L. et al. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12, 402 (2011). https://doi.org/10.1186/1471-2164-12-402

Aykut B, Pushalkar S, Chen R et al. The fungal mycobiome promotes pancreaticoncogenesis via activation of MBL. Nature. 2019;574:264–7.

Dierckxsens N, Mardulyn P, Smits G. NOVO Plasty: assembly of organelle genomes from whole genome data. Nucleic acids research. 2017;45(4):e18-e18.

Findley K, Oh J, Yang J et al. Topographic diversity of fungal and bacterialcommunities in human skin. Nature. 2020;498:367–70.

Hirai A, Kano R, Makimura K, Duarte ER, Hamdan JS, Lachance MA, Yamaguchi H,HasegawaA. Malassezia nana sp. nov., a novel lipid-dependent yeast species isolated from animals. Int J Syst Evol Microbiol. 2004 Mar;54(Pt 2):623-627. doi: 10.1099/ijs.0.02776-0. PMID: 15023986.

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. Model Finder: fast model selection for accurate phylogenetic estimates. Nature Methods 2017;4(6):587-589.

Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013;30(4):772– 780.

Limon JJ, Tang J, Li D et al. Malassezia is associated with Crohn's disease andexacerbates colitis in mouse models. Cell Host Microbe. 2019;25:377–88.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQTREE 2: new models and ef cient methods for phylogenetic
inference in the genomic era. Molecular Biology and Evolution 2020;37(5):1530-1534.

Saunte DML, Gaitanis G, Hay RJ. Malassezia-Associated skin diseases, the useof diagnostics and treatment. Front Cell Infect Microbiol. 2020;10:112.

Spatz M, Richard ML. Overview of the potential role of Malassezia in guthealth and disease. Front Cell Infect Microbiol. 2020;10:201.

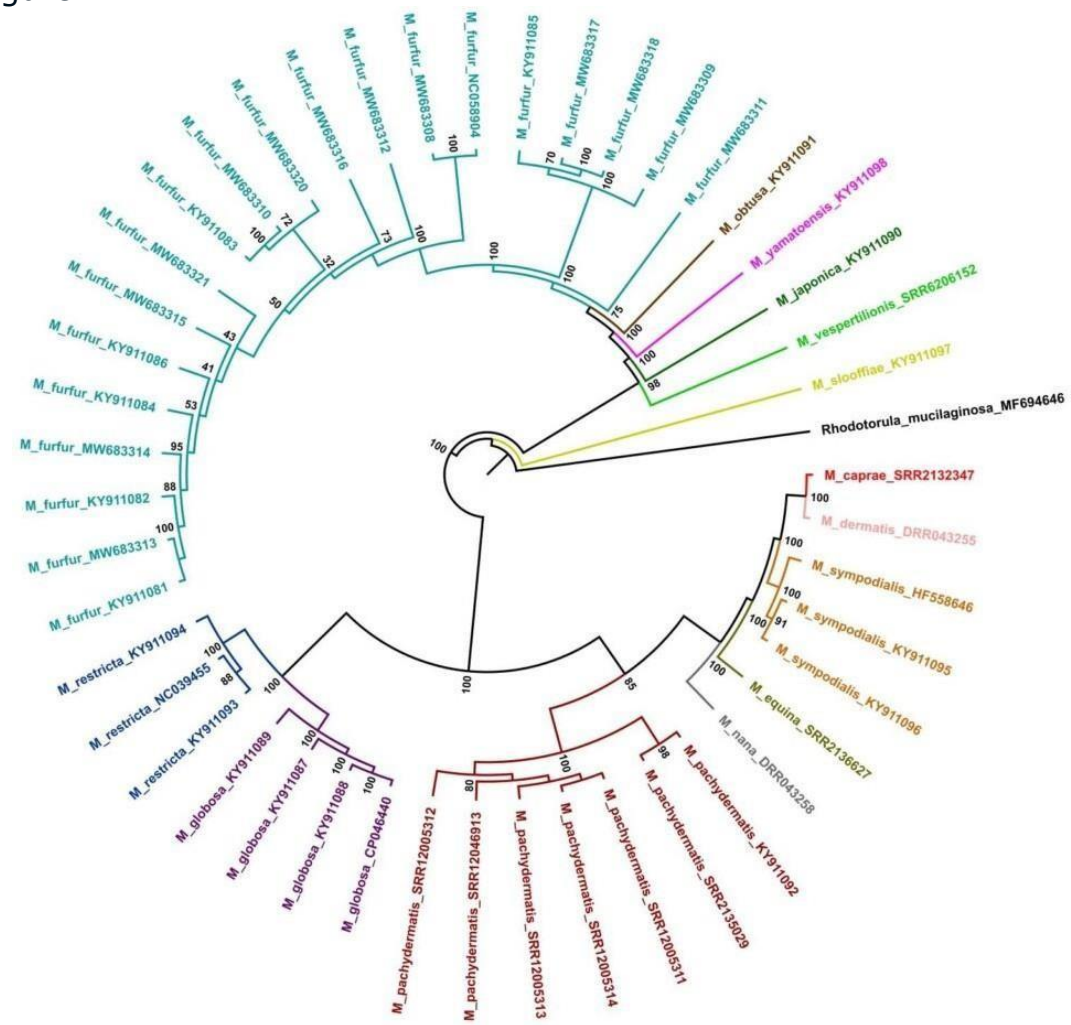
Theelen B, Cafarchia C, Gaitanis G et al. Malassezia ecology, pathophysiology,and treatment [published correction appears in Med Mycol. 2019 Apr 1;57(3):e2]. Med Mycol. 2018;56:S10– 25.

Vaidya G, Lohman DJ, Meier R. Sequence Matrix: concatenation software forthe fast assembly of multi-gene datasets with character set and codon information. Cladistics 2011;27(2):171-180.

Wang QM, Theelen B, Groenewald M, Bai FY, Boekhout T. Moniliellomycetes andMalasseziomycetes, two new classes in Ustilaginomycotina. Persoonia- Molecular Phylogeny and Evolution of Fungi. 2014;33(1):41-47.

Wu, G., Zhao, H., Li, C., Rajapakse, M. P., Wong, W. C. et al. Genus-wide comparative genomics of Malassezia delineates its phylogeny, physiology,and niche adaptation on human skin. PLoS Genetics 11, e1005614 (2015)

Figure 1.



---

[1] Integrated Biotechnology Center, University of Mogi das Cruzes, Mogi das

Cruzes, SP, Brazil

[2] Technological Research Center, University of Mogi das Cruzes, Mogi das Cruzes,

SP, Brazil

[3] Laboratory of Ecological and Evolutionary Genetics, Federal University of Viçosa, Rio Paranaíba, MG, Brazil.

RevistaFT

A RevistaFT é uma Revista Cientí ca Eletrônica Multidisciplinar Indexada de Alto Impacto e Qualis "B2" em 2023. Periodicidade mensal e de acesso livre. Leia gratuitamente todos os artigos e publique o seu também clicando aqui.

Contato

Queremos te ouvir.
WhatsApp: 21 98159-7352
e-Mail: contato@revistaft.com.br
ISSN: 1678-0817
CNPJ: 48.728.404/0001-22

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), fundação do Ministério da Educação (MEC), desempenha papel fundamental na expansão e consolidação da pós-graduação stricto sensu (mestrado e doutorado) em todos os estados da Federação.

Conselho Editorial

Editores Fundadores:

Dr. Oston de Lacerda Mendes.

Dr. João Marcelo Gigliotti.

Editor Cientí co:

Dr. Oston de Lacerda Mendes

Orientadoras:

Dra. Hevellyn

Andrade Monteiro

Dra. Chimene Kuhn

Nobre

Revisores:

Lista atualizada periodicamente em revistaft.com.br/expediente

Venha fazer partede nosso time de revisores também!

Copyright © Editora Oston Ltda. 1996 - 2023

Rua José Linhares, 134 - Leblon | Rio de Janeiro-RJ | Brasil

Data in Science
LADS

IDENTIFICATION AND PROFILING

Authors

Gledison Eric Teixeira[1]

Gabriel Barbosa Huszcz[1,2]

David Aciole Barbosa[1]

Igor Henrique Rodrigues-

Oliveira[3] Renan Rodrigues

Rocha[2,3]

Karine Frehner Kavalco[3]

Rubens Pasa[3]

Fabiano Bezerra Menegidio[1,2]

Affiliation

1. Integrated Biotechnology Center, University of Mogi das Cruzes, Mogidas Cruzes, SP, Brazil.
2. Technological Research Center, University of Mogi das Cruzes, Mogi das Cruzes, SP, Brazil.
3. Laboratory of Ecological and Evolutionary Genetics, Federal Universityof Viçosa, Rio Paranaíba, MG, Brazil.

Correspondence Author

Fabiano Bezerra Menegidio (fabianomenegidio @umc.br)

Prior publications

Huszcz G.B., Teixeira G.E., Rodrigues-Oliveira I.H., Rocha R.R., Oliveira R.C., Takao P.S., Barbosa D.A., Jabes D.L., Kavalco K.F., Pasa R., Menegidio F.B. **Five mitochondrial genomes of the lipophilic fungus *Malassezia* (*Malasseziales*, *Exobasidiomycetes*)**. RevistaFT, Edição 124 JUL/23 SUMÁRIO / 12/07/2023. DOI: 10.5281/zenodo.8140802

Acknowledgements

This study was financed in part by scholarship grants from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil - CAPES (www.capes.gov.br/) and Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (www.gov.br/cnpq/pt-br).

Reviewers

- Bruno Cavalhero Araújo, bru.biol1@gmail.com, Cawthron Institute.
- Daniela Leite Jabes, danielajabes@umc.br, Universidade de Mogi das Cruzes.
- Caio Augusto Perazza, caioperazza@gmail.com, UniversidadeEstadual Paulista.
- Micheli Zaminhan-Hassemer, michelizam@hotmail.com,Universidade de Mogi das Cruzes.

Subject area

Biological Sciences; Microbiology

Data Access

Inform where the data can be accessed:
( ) Together with the article: when the data are submitted with the articleand will be accessible in LADS itself.
(x) In a qualified institutional repository: LADS suggests depositing in Mendeley Data. Other repositories will be evaluated.

In case of a repository, please inform
Repository name: Open Science Framework (OSF)
DOI of the dataset (when available): 10.17605/OSF.IO/QMZKS
Link to access the data: https://osf.io/qmzks/

Social Networks

@labiomics (Instagram and Twitter)

ARTICLE STRUCTURE

Article Title

Mining mitochondrial genomes of *Malassezia* fungi: building a comprehensive dataset and database

Abstract

The study of mitochondrial genomes in *Malassezia* fungi holds significant importance in understanding their genetic diversity and functional characteristics. In this research, we present a comprehensive approach to mining and analyzing mitochondrial genomes of various *Malassezia* species. Our primary objective was to build a comprehensive dataset that encompasses a wide range of genomic information from these fungi. Leveraging this dataset, we implemented a relational database that serves as a powerful resource for storing, managing, and querying mitochondrial genomic data. Through a combination of bioinformatic tools, we conducted functional annotation of the mitochondrial genomes, identifying crucial genes, regulatory elements, and potential functional pathways. This annotation process was incorporated into the database, enhancing its utility for researchers and facilitating cross-species comparisons. The constructed dataset and database provide an invaluable resource for researchers studying *Malassezia* fungi and their mitochondrial genomes. The database's user-friendly interface allows for easy access to genomic data, enabling efficient exploration and analysis. Additionally, the dataset opens avenues for in-depth investigations into the evolution and adaptation of these fungi to diverse ecological niches. This article highlights the significance of a comprehensive dataset and database in the field of mitochondrial genomics and its potential to advance our knowledge of *Malassezia* fungi. The integration of functional annotations with the genomic data will foster new discoveries and contribute to broader studies on fungal biology and evolution.

Keywords

Malassezia; mitogenomes; next-generation sequencing; relatio nal database; Machado database

Data importance

- *Malassezia* fungi comprise a diverse group of species that inhabit diverse environments, including the skin of humans and other animals (Findley et al., 2020). Some members of this genus have been linked to skin disorders, while others play crucial roles in industrial processes, such as lipid biotransformation (Theelen *et al.*, 2018; Aykut et al., 2019; Limon et al., 2019; Saunte *et al.*, 2020; Spatz and Richard 2020). However, despite their significance, a comprehensive understanding of the mitochondrial genomes of *Malassezia* fungi is still lacking.
- The primary objective of this study is threefold: to perform functional annotation of mitochondrial genomes in a variety of *Malassezia* fungal species, unveiling their genetic features and functional elements. This comprehensive dataset, comprising twenty *Malassezia* mitogenomes, was compiled through state-of- the-art high-throughput sequencing and extensive bioinformatic analysis. The resulting insights into vital mitochondrial genes, regulatory regions, and functional pathways shed light on their adaptation and biological roles in these fungi.
- Secondly, we have successfully constructed a relational database based on the project Machado (Mudadu and Zerlotini, 2020), utilizing the Chado schema within the PostgreSQL SGDB (Mungall and Emmert, 2007). This database server as a centralized repository for researchers interested in exploring and analyzing mitochondrial genomes of *Malassezia* fungi, facilitating cross-species comparisons, and enabling future investigations. With the database readily available, researchers can efficiently access and utilize the data to delve deeper into the genetic intricacies of these fungi, accelerating the pace of scientific discoveries and advancements in various fields of fungal genomics. The adoption of the Chado schema ensures a robust and standardize d structure, enhancing the database's utility and interoperability within the scientific community.
- Lastly, as a significant achievement, we have successfully developed a web application based on the Sequence Server framework (Priyamet al., 2019), integrated with the service of containers Docker (https://www.docker.com/). This user-friendly web application provides an intuitive interface for researchers to access and explore the comprehensive genomic data stored in our database. Leveraging the power of Docker containers ensures easy deployment and scalability, allowing seamless access to the application across various platforms and environments.

The implementation of this web application represents a major step forward in enhancing the accessibility and usability of the genomic data, empowering researchers to delve deeper into the genetic intricacies of these fungi and further accelerating scientific discoveries in the field of fungal genomics.

Material and Methods

Data Collection and Preprocessing

For this work, we obtained a total of 40 mitochondrial genomes belonging to 10 species of fungi from the genus *Malassezia*. These genomes were retrieved from the Nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore) and/or the Organelle Genome Resources (https://www.ncbi.nlm.nih.gov/genome/organelle/) maintained by the National Center for Biotechnology Information (NCBI). Additionally, during the compilation of the dataset and database, we incorporated 5 mitochondrial genomes presented in the work of Huszcz et al. (2023).

To ensure the accuracy and integrity of the downloaded files, a thorough data integrity check was carried out. This step aimed to detect any corruption or issues that might have occurred during the download process.

All relevant metadata associated with the mitochondrial genomes, including species information, DNA sequences, annotations, and other details, were then extracted from the downloaded files and meticulous ly documented for future reference.

To maintain consistency in the structure of the data, the obtained mitochondrial genomes were standardized. This involved ensuring uniformity in file formats and sequence nomenclature across all genomes.

Functional Annotation

The functional annotation of the mitochondrial genomes was performed using Mitos2 (http://mitos2.bioinf.uni-leipzig.de/; Arab et al., 2017) in conjunction with Gene Code 3 - Yeast and RefSeq 89 Fungi databases. This annotation process aimed to identify genes encoding essential mitochondrial proteins, regulatory regions, and potential functional pathways within the genomes.

Implementation of the Machado Framework

In this study, we successfully implemented the Machado Framework (Mudadu and Zerlotini, 2020) on a server to establish a robust infrastructure for the database. The implementation process comprised several crucial steps, adhering to instructions available in the research article and

the project's GitHub repository.

To begin, we carefully followed the official Docker instructions to install Docker infrastructure (https://docs.docker.com/engine/install/ubuntu/) and Docker Compose tool (https://docs.docker.com/compose/) on the server. These tools are essential for the smooth operation of the Machado Framework and the efficient execution of the required containers.

Furthermore, to optimize container construction and minimize image size, we performed a refactoring of the provided Dockerfile. By reducing the use of the RUN parameter, we minimized the number of layers generated during container construction. Subsequently, we created a new Docker Compose file, which referred to our updated Machado Framework image.

With the server ready and the Docker containers properly configured, we proceeded to implement the Machado Framework environment by cloning the Machado Framework repository. As part of this setup, we created directories for PostgreSQL (https://www.postgresql.org/), ElasticSearch (https://www.elastic.co/what-is/elasticsearch), and JBrowse (Diesh et al., 2023) data, and carefully edited the Machado environment variables in the .env file.

Upon replacing the original Dockerfile and docker-compose.yml files with our refactor red versions, we executed the Machado environment using Docker Compose. This deployment enabled convenient access to the web interface of the Machado Framework athttp://hostname:8000/machadosample/.

Overall, this comprehensive implementation allowed us to offer a powerful and user-friendly platform, empowering researchers to explore and analyze mitochondrial genomes of various *Malassezia* fungal species. This significant advancement fosters progress in the field of fungal genomics and holds great potential for broader applications in various scientific domains.

Implementation of the Sequence Server

To facilitate access to the mitochondrial genomes, we provided a database based on the Sequence Server (Priyam et al., 2019) that allowed the search for sequence similarity through the BLAST tool (Altschul et al., 1990). This tool displays alignment results in visual graphics which can be downloaded in SVG and PNG formats. It also provides direct download of sequence and alignment from the result page.

Data description

Malassezia mitogenome dataset

Resulting from this research, a comprehensive General Feature File (GFF3) of ~13 kilobytes has been generated and is available for retrieval from the Open Science Framework (OSF) repository (see supplementary materials). This GFF3 file contains valuable data concerning protein coding genes (PCGs), transfer RNAs (tRNAs), ribosomal RNA (rRNAs), mtDNA control region (D-loop), and other elements present in *Malassezia* mitogenomes. The information contained in this file holds significance across various fields, including taxonomy, ecology, physiology, and evolution. The GFF3 file's annotation consists of nine columns, each serving a specific purpose:

1. Identity: This column provides the name of the scaffold where the annotated sequence is located.
2. Method: Here, the algorithm or operating procedure responsible for generating the sequence is described, such as “mitos” or “mitfi”.
3. Type: This column denotes the name of the annotated resource type, which could be “gene”, “exon”, “mRNA”, “tRNA”, “rRNA” or “ncRNA_gene”.
4. Start: Indicates the position from which the genomic resource is noted within the scaffold, with a displacement of one base.
5. End: Represents the end position of the annotated genomic resource, with an offset of one base.
6. Score: This value indicates the confidence level of the source in the annotated resource. When the confidence level is indeterminate, a dot “.” is used.
7. Strand: The presence of a single character indicates the strand of the annotated resource. “+” sign represents the positive or 5'->3' strand, “-” denotes the negative or 3'->5' strand, and “.” indicates an undetermined strand.
8. Phase: This column uses values 0, 1, or 2 to indicate the resource phase of the coding sequence (CDS), while “.” is used for other types of resources.
9. Attributes: This column contains all other relevant information related to the annotated resource.

In addition to the consolidated GFF3 file, we are pleased to inform that we have also made available the annotations of the 45 mitogenomes in two additional formats: BED-6 and TBL. These annotations offer further insights into the genomic features of the *Malassezia* mitogenomes. The BED-

6 format provides a concise representation of the annotation data and includes the following six columns:

1. Chromosome/Identity: The name of the scaffold where the annotated sequence is located.
2. Start: The position from which the genomic resource is noted within the scaffold, with a displacement of one base.
3. End: The end position of the annotated genomic resource, with anoffset of one base.
4. Name: A label or identifier for the annotated resource.
5. Score: This value indicates the confidence level of the source in the annotated resource. When the confidence level is indeterminate, a dot "." is used.
6. Strand: The presence of a single character indicates the strand of theannotated resource. "+" sign represents the positive or 5'->3' strand, "-" denotes the negative or 3'->5' strand, and "." indicates an undetermined strand.

On the other hand, the TBL format provides a tab-separated table that includes more detailed information about the annotation data, enhancing the understanding of the genomic features. It encompasses various columns that capture relevant details about each annotated resource, such as start and end positions, strand orientation, attributes, and more.

We have also made available the data in the FASTA format. These FASTA files contain the complete mitochondrial genomes of the organisms,along with the nucleotide and protein sequences of protein- coding genes,tRNAs, rRNAs, and other elements present in the *Malassezia* mitogenomes.The FASTA formatis widely used in bioinformatics and genomics research, as it provides a simple and efficient way to represent nucleotideand protein sequences. Each sequence in the FASTA file is denoted by aheader line, starting with a ">" symbol, followed by a description or identifier for the sequence. The actual sequence data follows the header line.

The availability of the complete mitochondrial genomes in the FASTA format allows researchers to conduct in-depth analyses, such as sequence alignments, phylogenetic studies, and comparative genomics. Moreover, the individual sequences of protein-coding genes, tRNAs, rRNAs, and other elements offer valuable insights into the functional and structural aspects of the *Malassezia* mitogenomes.

Implementation of *Malassezia*DB

To ensure easy access and utilization of the valuable dataset, we established a robust infrastructure using the Machado Framework (Mudadu and Zerlotini, 2020) for the database. The entire dataset, along with necessary scripts and a Docker Compose file, is available for retrieval from the Open Science Framework (OSF) repository.

Efficient population of the database tables is facilitated by a custom script that handles data from FASTA and GFF files. This script is provided in the OSF repository, allowing researchers to query and analyze genomic information effortlessly. To support local execution of the database, we included a Docker Compose file in the OSF repository. Researchers can easily initiate an instance of the database and perform in-depth analyses on *Malassezia* mitochondrial genomes without complex setup procedures.

The user-friendly interface of the Mitogenome Database (Figure 1) enables users to interact with and explore available mitogenomes with intuitive navigation and search functionalities.

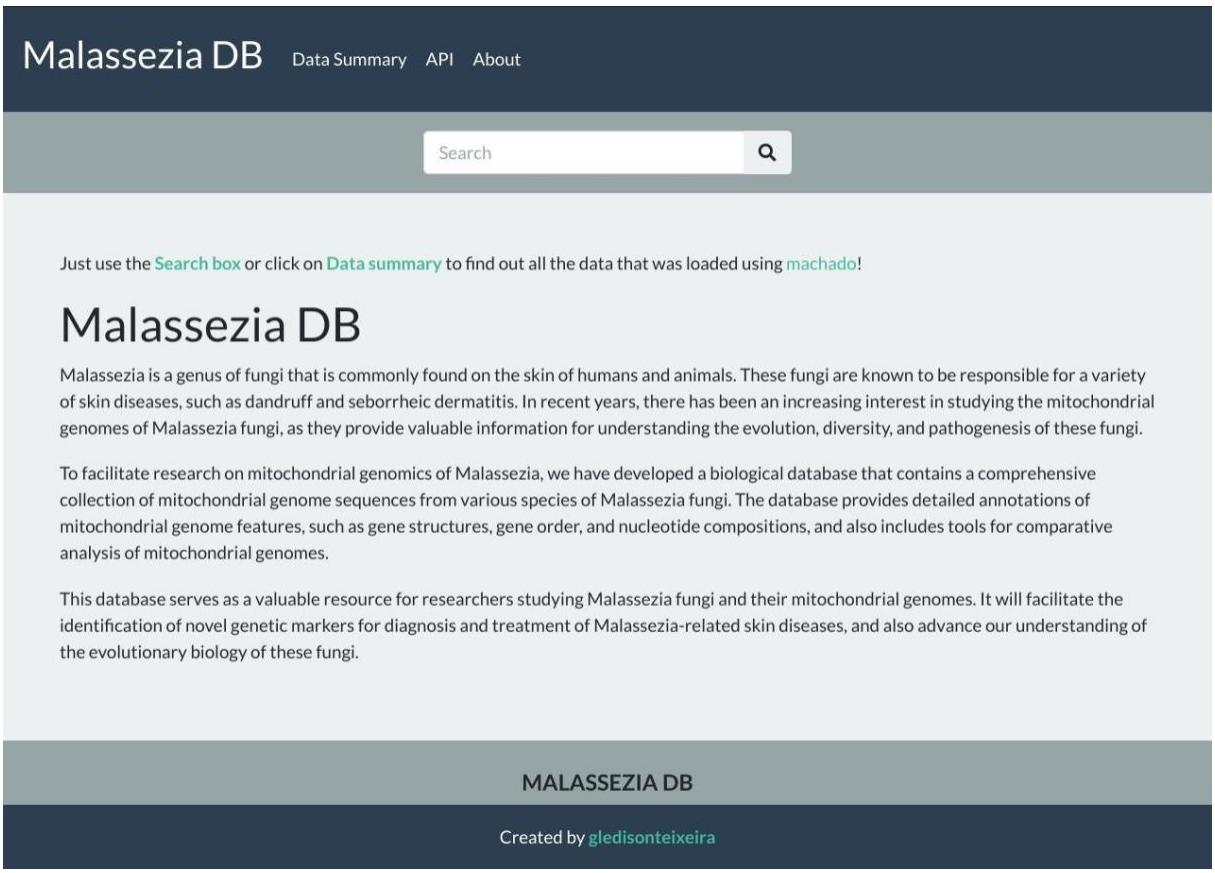


Figure 1. Interface of the Mitogenome Database

The research and display interface of the database (Figure 2) allows users to execute sophisticated queries, visualize mitochondrial genomes, and access detailed annotation information, providing valuable insights into genomic features.

The Genome Browser is a complementary tool that enables researchers to visualize the positional information of genes and genetic elements within *Malassezia* mitochondrial genomes in a graphical format. It offers an interactive and dynamic display of annotated features, allowing users to explore genomic organization and arrangement easily.

The Genome Browser enhances the exploration and interpretation of genetic elements, helping researchers identify potential regulatory regions, conserved motifs, and structural variations. It aids in the identification of gene clusters and overlaps, contributing to functional relationship and evolutionary insights.

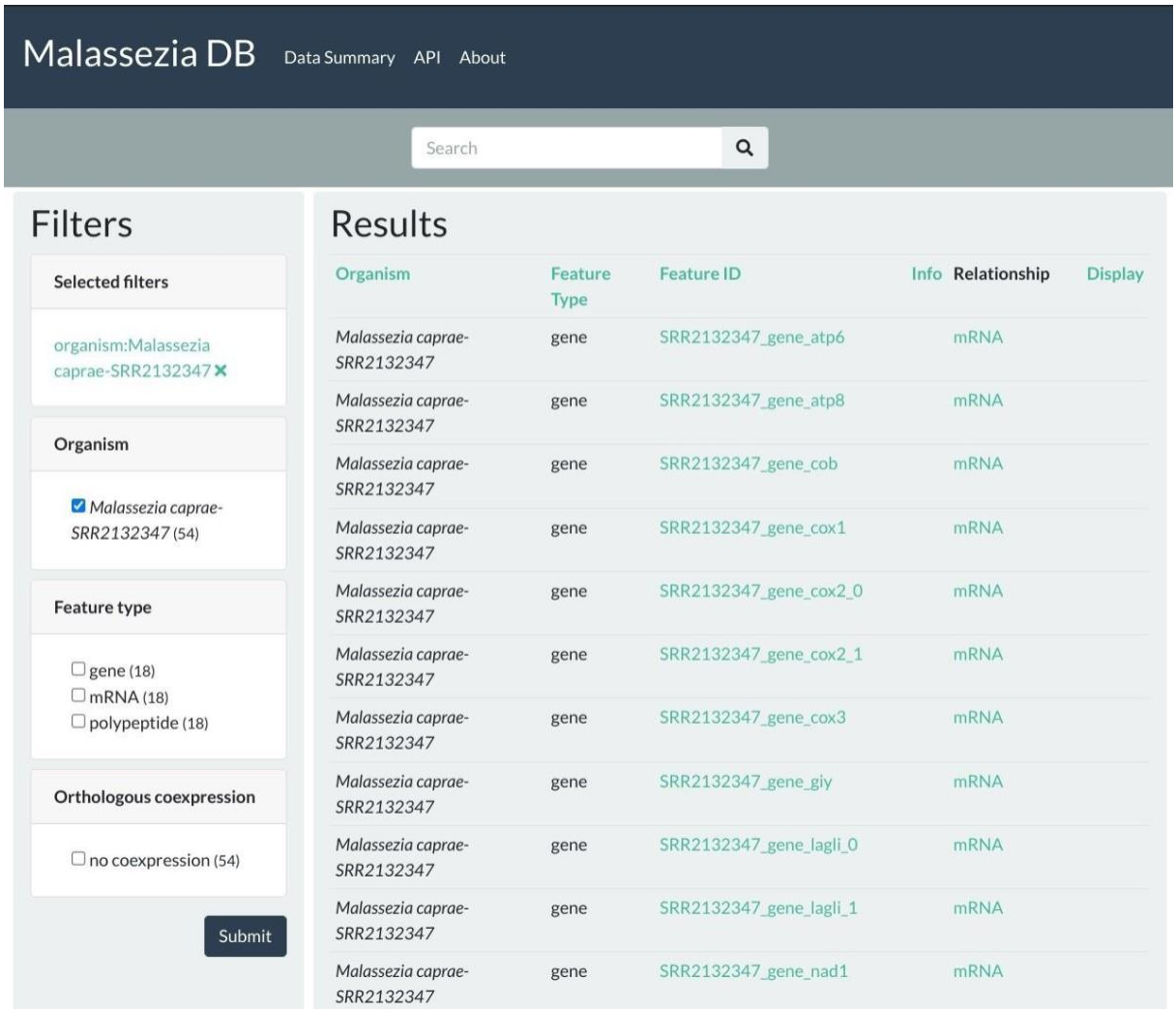


Figure 2. Search and display interface of the Database

Figure 3 provides a glimpse of the Genome Browser interface, showcasing its capability to empower researchers in understanding the genomic landscape of *Malassezia* organisms.

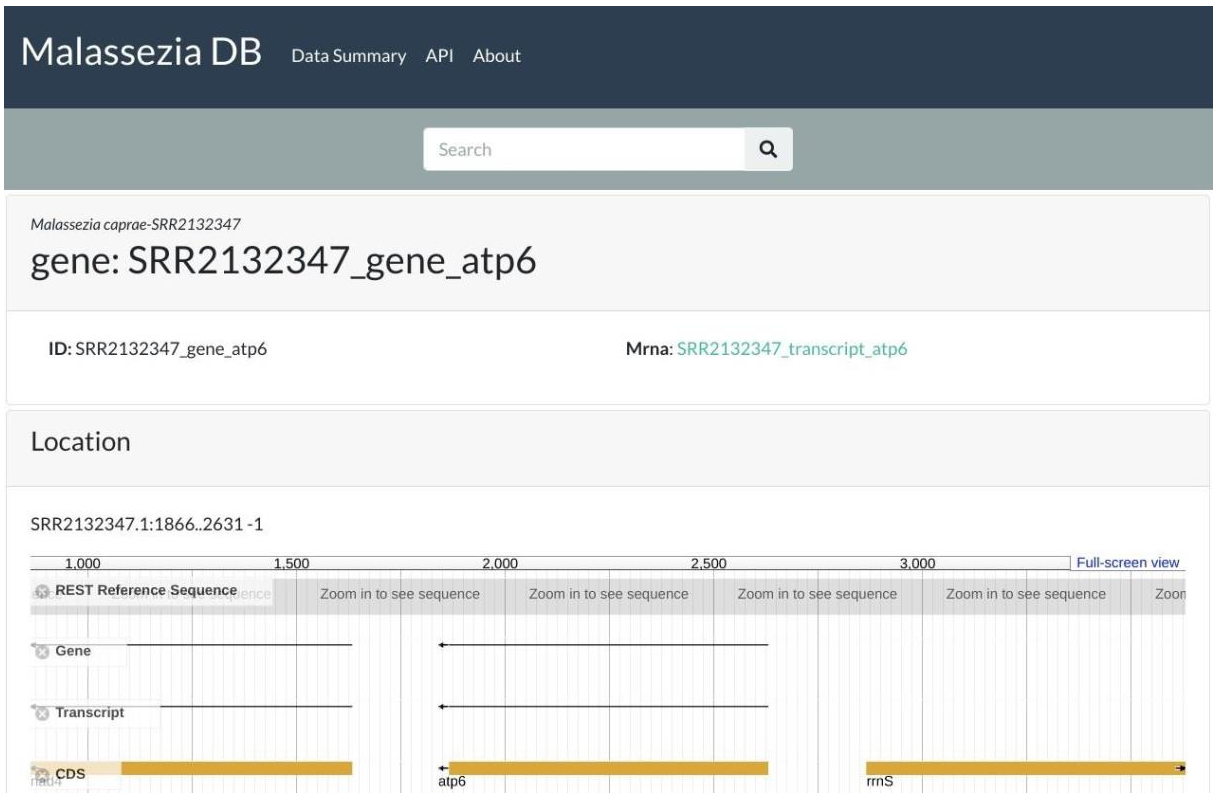


Figure 3. Genome Browser

The Chado schema implemented in *Malassezia*DB (Figure 4) servesas the backbone for organizing and structuring genomic data. This well- established and widely used relational database schema ensures efficient and coherent data retrieval and management, fostering collaboration among researchers.

The Chado schema's robustness and versatility within *Malassezia*DB support advanced genomic analyses and interpretation, driving advancements in fungal genomics and facilitating discoveries that benefit multiple scientific disciplines.

Overall, the integrated *Malassezia*DB platform presents a powerful resource for researchers to conduct comprehensive and visually intuitive investigations into the mitochondrial genomes of *Malassezia* organisms. We encourage users to leverage these tools and resources for furthering their research endeavors in the field of fungal genomics.

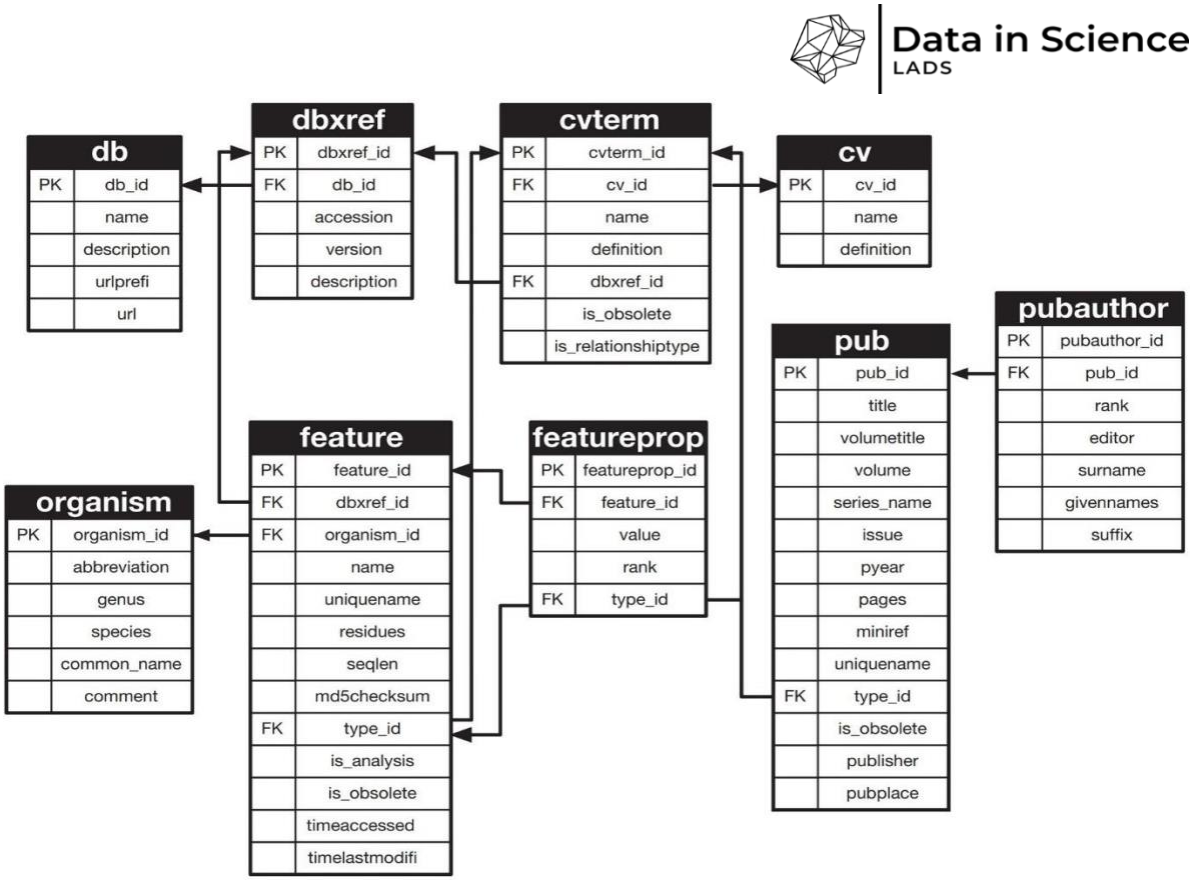


Figure 4. Chado Schema

Implementation of *Malassezia*BLAST

*Malassezia*BLAST offers researchers the unique capability to conduct comprehensive comparisons and analyses of gene and protein sequences derived from the 45 mitochondrial genomes of 10 distinct *Malassezia* species, all of which are available in the comprehensive *Malassezia* Dataset and *Malassezia*DB. Seamlessly integrated with Sequence Server (Priyam et al., 2019), this cutting-edge platform facilitates sequence alignment and exploration, providing valuable insights into the genetic composition of these fungal organisms.

To initiate sequence alignment, users can conveniently input gene sequences or import sequence files directly into the designated input box (Figure 5). For a personalized approach, researchers can fine-tun e parameters, such as -e-value, -num_alignments, and more, through the option box, ensuring tailored and precise results. The BLAST output delivers comprehensive alignment results, including the queried sequence and corresponding references from the database, facilitating the identification of homologous genes or specific genes of interest (Figure 6).

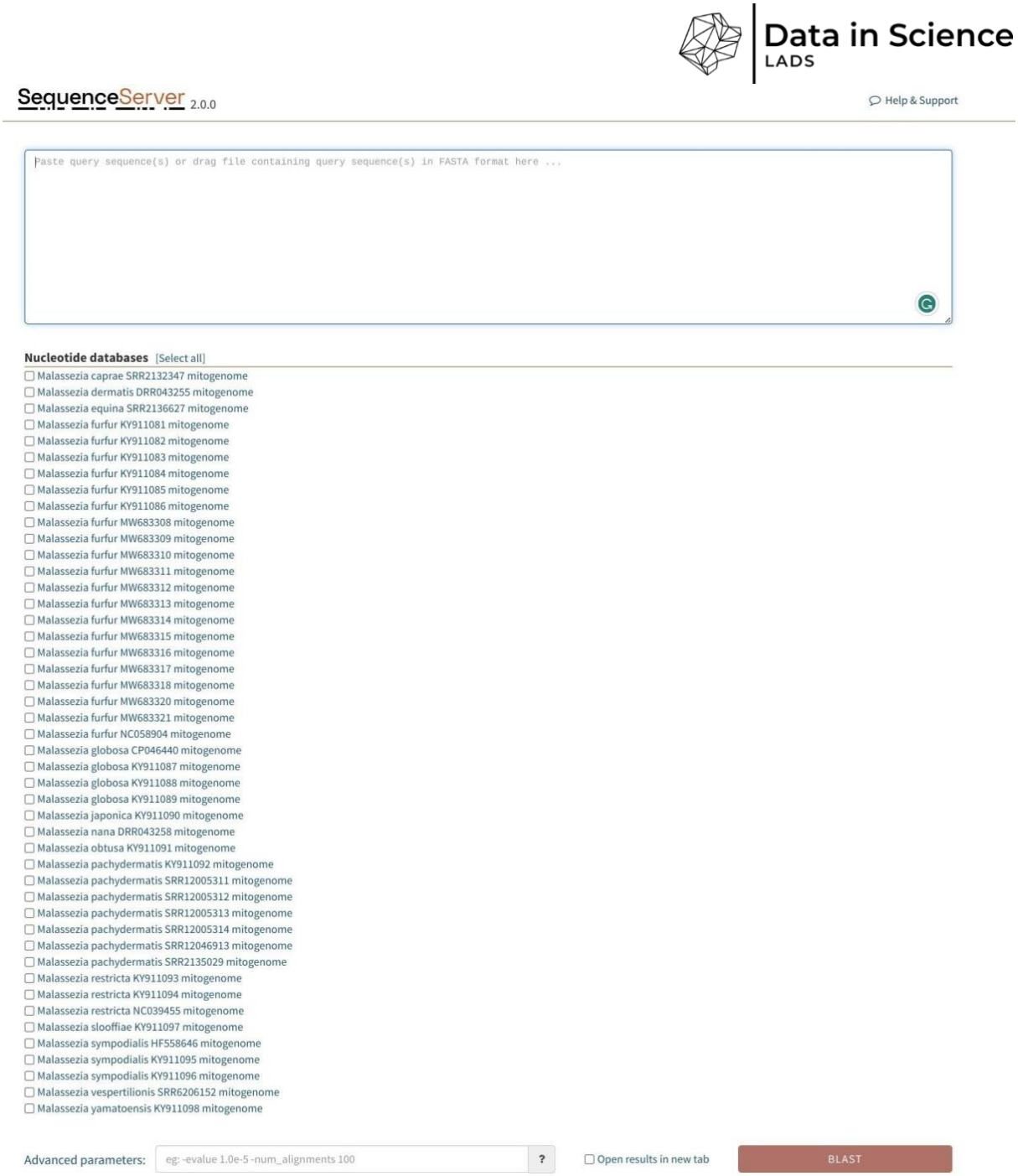


Figure 5. Interface of the *Malassezia*BLAST

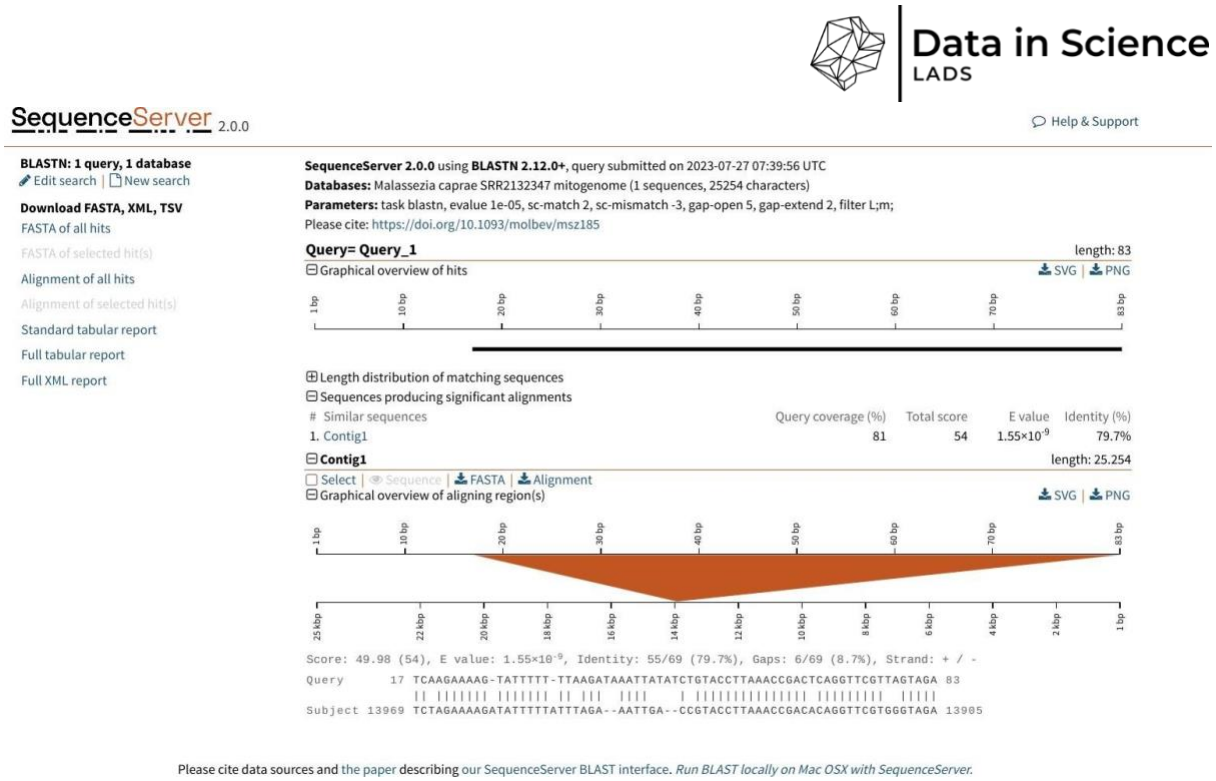


Figure 6. Similarity search result in *Malassezia*BLAST interface

References

- Al Arab M, Zu Siederdissen CH, Tout K, Sahyoun AH, Stadler PF, Bernt M. Accurate annotation of protein-coding genes in mitochondria l genomes. Molecular phylogenetics and evolution. 2017;106: 209-216.
- Altschul, S. F., et al. Basic Local Alignment Search Tool. Journal of Molecular Biology 215, 403–410 (1990) doi:10.1016/S0022-2836(05)80360-2
- Aykut B, Pushalkar S, Chen R et al. The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. Nature. 2019;574:264–7.
- Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, Guo E, Dider S, Zhang J, Bridge C, Hogue G, Duncan A, Morgan M, Flores T, Bimber BN, Haw R, Cain S, Buels RM, Stein LD, Holmes IH. JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol. 2023 Apr 17;24(1):74. doi: 10.1186/s13059-023-02914-z. PMID: 37069644; PMCID: PMC10108523.
- Findley K, Oh J, Yang J et al. Topographic diversity of fungal andbacterial communities in human skin. Nature. 2020;498:367–70.
- Huszcz G.B; Teixeira G.E; Rodrigues-Oliveira I.H; Rocha R.R; Oliveira RC; Takao PS; Aciole Barbosa D.; Jabes DL; Kavalco KF; Pasa R; Menegidio F.B. Five mitochondrial genomes of the lipophilic fungus *Malassezia* (*Malasseziales*, Exobasidiomycetes). RevistaFT, v. 27, n.124, p. 1-2, 2023.
- Limon JJ, Tang J, Li D et al. *Malassezia* is associated with Crohn's disease and exacerbates colitis in mouse models. Cell Host Microbe. 2019;25:377–88.
- Mudadu MA, Zerlotini A. Machado: Open source genomics data integration framework. Gigascience. 2020 Sep 14;9(9):giaa 097. doi: 10.1093/gigascience/giaa097. PMID: 32930331; PMCID: PMC7490629.
- Mungall C.J., Emmert D.B., FlyBase Consortium. (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. Bioinformatics, 23, i337–i346.
- Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, Chowdhary H, Pieniak I, Maynard LJ, Gibbins MA, Moon H, Davis- Richardson A, Uludag M, Watson-Haigh NS, Challis R, Nakamura H, Favreau E, Gómez EA, Pluskal T, Leonard G, Rumpf W, Wurm Y. Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. Mol Biol Evol. 2019 Dec 1;36(12):2922 -2924. doi: 10.1093/molbev/msz185. PMID: 31411700; PMCID: PMC6878946.
- Saunte DML, Gaitanis G, Hay RJ. *Malassezia*-Associated skin diseases, the use of diagnostics and treatment. Front Cell Infect Microbio l. 2020;10:112.
- Spatz M, Richard ML. Overview of the potential role of *Malassezia* in gut health and disease. Front Cell Infect Microbiol. 2020;10:201.

· Theelen B, Cafarchia C, Gaitanis G et al. *Malassezia* ecology, pathophysiology, and treatment [published correction appears in Med Mycol. 2019 Apr 1;57(3):e2]. Med Mycol. 2018;56:S10–25.