





Universidade do Minho
Escola de Ciências

Cristiana Sofia Codeço Leal

Unravelling the genomic and metabolic features of the yeast *Cyberlindnera jadinii*: towards the improvement of microbial cell factories

Dissertação de Mestrado

Mestrado em Bioquímica Aplicada

Trabalho realizado sob a orientação de

Doutora Isabel João Soares Silva

Doutor Ricardo Filipe Azevedo Franco Duarte

Dezembro 2023

Despacho RT - 31 /2019 - Anexo 3

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada. Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.



Atribuição-NãoComercial-SemDerivações

CC BY-NC-ND

AGRADECIMENTOS

Chega ao fim uma etapa muito importante do meu percurso académico e apesar de não ter sido uma fase fácil nem como esperaria tenho muito a agradecer.

Em primeiro lugar à minha orientadora, Doutora Isabel João, pela disponibilidade, compreensão, partilha de conhecimento e rigor científico, ajuda nas correções de última hora e também pela simpatia. Ao Doutor Ricardo Duarte, imprescindível na execução e análise da componente bioinformática deste trabalho, obrigada pela ajuda, disponibilidade e rigor.

A toda a equipa do laboratório LBM II, obrigada pela partilha de conhecimentos, em particular ao João, o meu maior agradecimento por me ensinar tanto e ajudar na maior parte das tarefas. À Tayane um agradecimento especial pela simpatia e disponibilidade em me ajudar sempre que necessário.

Ao Pedro, a pessoa que me apoia sempre e não me deixou desistir, um pilar fundamental tanto na vida académica como pessoal. Obrigada pela força, paciência, companheirismo, motivação. A ti os agradecimentos nunca serão suficientes.

Aos meus amigos da licenciatura e às meninas que conheci no primeiro ano de universidade, obrigada pela amizade e pelos momentos e memórias que ficarão para sempre.

E por último, os mais importantes. Os que apoiam incondicionalmente. Obrigada pais, irmãs, avó e madrinha pela preocupação, incentivo e pelas palavras que aconchegam.

Despacho RT - 31 /2019 - Anexo 4

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho acadêmico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

**CARATERIZAÇÃO DO GENOMA E DO METABOLISMO DA LEVEDURA *CYBERLINDNERA JADINII* PARA A
OTIMIZAÇÃO DE FÁBRICAS CELULARES MICROBIANAS**

RESUMO

O desenvolvimento de aplicações industriais utilizando sistemas biológicos tem ganho uma importância crescente, como uma alternativa aos métodos de síntese química, na procura de processos mais inovadores e sustentáveis. As leveduras, essenciais em vários processos industriais, têm sido alvo de estudo dada a sua capacidade de adaptação às condições adversas presentes em vários destes processos, assim como a capacidade de produzir vários metabolitos de interesse. A levedura *Cyberlindnera jadinii* tem atraído o interesse de vários investigadores dado o seu elevado potencial biotecnológico. Esta levedura possui uma grande robustez, tem capacidade para crescer em meios de cultura baratos apresentando elevados níveis de aplicabilidade e de produtividade. É também capaz de utilizar uma gama alargada de ácidos orgânicos como fonte única de carbono e energia. Nesse sentido, o estudo da sua fisiologia, metabolismo e genoma é de extrema relevância para potenciar o seu desenvolvimento como fábrica celular microbiana.

Neste trabalho, analisámos seis estirpes de *C. jadinii*: CBS 1600, DSM 2361, PYCC 2578, PYCC 3092, TB104 e TB115. Inicialmente avaliámos o seu perfil de utilização de ácidos carboxílicos como única fonte de carbono e de energia. Com o objetivo de identificar novos transportadores de ácidos carboxílicos de *C. jadinii* efetuámos a expressão heteróloga de transportadores putativos em *Saccharomyces cerevisiae*, tendo em vista a caracterização funcional. Em paralelo, numa abordagem *in silico*, explorámos o genoma desta levedura através de uma análise genómica comparativa entre cinco estirpes. Todas as estirpes de *C. jadinii* analisadas foram capazes de crescer eficientemente em meios contendo ácidos mono e dicarboxílicos, como única fonte de carbono e energia, mas no ácido tricarboxílico (ácido cítrico) o crescimento foi dependente da estirpe e da temperatura de crescimento. A expressão do gene CT5 em *S. cerevisiae* promoveu um decréscimo de crescimento em todos os meios testados, o que sugere que este pode codificar um exportador de ácidos carboxílicos. Relativamente à análise genómica, as estirpes de *C. jadinii* apresentam variabilidade genética, incluindo diferentes níveis de ploidia e presença de inserções e deleções quando comparadas com o genoma de referência. A estirpe PYCC 2578 é a mais distinta geneticamente. Este trabalho contribuiu para uma compreensão mais profunda da levedura *C. jadinii*, destacando o seu potencial biotecnológico. **Palavras-chave:** ácidos carboxílicos; análise genómica; *Cyberlindnera jadinii*; expressão heteróloga

**UNRAVELLING THE GENOMIC AND METABOLIC FEATURES OF THE YEAST *CYBERLINDNERA JADINII*:
TOWARDS THE IMPROVEMENT OF MICROBIAL CELL FACTORIES**

ABSTRACT

The strong demand for industrial applications using biological systems has become increasingly important as an alternative to conventional methods given the growing need for innovative and sustainable solutions. Yeasts, which are essential in industrial processes, have been explored to understand their adaptability to adverse fermentation conditions, and ability to produce metabolites of interest. The yeast *Cyberlindnera jadinii* has attracted the interest of researchers due to its high biotechnological potential. This yeast is very robust and has the capacity to grow in inexpensive culture media, showing high levels of applicability and productivity. It can also use a wide range of organic acids as a sole source of carbon and energy source. In this regard, exploring its physiological, metabolic, and genomic characteristics is extremely important to promote its implementation as a microbial cell factory.

In this study, we analyzed six strains of *C. jadinii*: CBS 1600, DSM 2361, PYCC 2578, PYCC 3092, TB104 e TB115. Initially we determined the assimilation profiles of carboxylic acids as the sole source of carbon and energy source. Aiming at the identification of novel carboxylic acid transporters, we heterologously expressed putative *C. jadinii* transporters in *Saccharomyces cerevisiae* for functional analysis. In parallel, in an in-silico approach, we carried out a comparative genomic analysis of 5 strains. All the strains were able to grow efficiently on media containing mono- and dicarboxylic acids, as sole carbon and energy source, but growth on tricarboxylic acid (citric acid) was strain and growth temperature dependent. The expression of the CT5 gene in *S. cerevisiae* promoted a decreased growth in all tested media, which suggests that it can encode a as carboxylic acid exporter. Regarding the genomic analysis, the *C. jadinii* strains presented genetic variability, including different ploidies, the presence of insertions and deletions, when compared to the genome of reference. The strain PYCC 2578 was found to be the most distinct genetically. Overall, this work contributed to a deeper understanding of the yeast *C. jadinii*, highlighting its potential in the biotechnology field.

Keywords: carboxylic acids; *Cyberlindnera jadinii*; genomic analysis; heterologous expression

ÍNDICE

AGRADECIMENTOS	iii
RESUMO	v
ABSTRACT	vi
Índice	vii
LISTA DE ABREVIATURAS E ACRÓNIMOS	ix
LISTA DE FIGURAS	x
LISTA DE TABELAS	xi
1. INTRODUÇÃO	1
1.1 A levedura <i>Cyberlindnera jadinii</i>	2
1.1.1 Fisiologia	2
1.1.2 A evolução da nomenclatura de <i>C. jadinii</i>	3
1.1.3 Relevância biotecnológica	4
1.1.4 Aplicações industriais de <i>C. jadinii</i>	6
1.2 Proteínas transportadoras de membrana	7
1.3 Estudo do genoma através da bioinformática.....	9
1.3.1 <i>Softwares</i> bioinformáticos.....	10
2. OBJETIVOS	16
3. MATERIAIS E MÉTODOS	17
3.1 Material biológico	17
3.2 Meios de cultura	17
3.3 Condições de crescimento.....	19
3.4 Preparação de células competentes de <i>Escherichia coli</i>	19
3.5 Extração de DNA genómico de <i>Cyberlindnera jadinii</i>	19
3.6 Extração de DNA plasmídico.....	20
3.7 Digestão de DNA plasmídico com enzimas de restrição	20
3.8 Purificação de DNA plasmídico de um gel de agarose	21

3.9	Amplificação do DNA e análise por eletroforese em gel de agarose	21
3.10	Clonagem dos genes de <i>C. jadinii</i> no vetor p416GPD	24
3.10.1	Confirmação da clonagem no plasmídeo p416GPD	24
3.10.2	PCR de colónia.....	25
3.10.3	Estratégias alternativas de clonagem.....	25
3.10.3.1	Extração de DNAP de um gel de agarose por excisão das bandas.....	25
3.10.3.2	Amplificação do plasmídeo p416GPD por PCR.....	26
3.10.4	Sequenciação dos genes de <i>C. jadinii</i> clonados no vetor p416GPD	27
3.11	Expressão heteróloga dos genes de <i>C. jadinii</i> em <i>S. cerevisiae</i>	27
3.11.1	Clonagem de genes de <i>C. jadinii</i> no vetor p416GPD pelo método de Gap Repair em <i>S. cerevisiae</i>	28
3.11.2	Extração de DNA plasmídico de levedura.....	28
3.12	Ensaio de crescimento em meio sólido (<i>drop-tests</i>)	28
3.13	Análise bioinformática do genoma de <i>C. jadinii</i>	29
4.	ANÁLISE E DISCUSSÃO DOS RESULTADOS.....	32
4.1	Amplificação de genes da levedura <i>Cyberlindnera jadinii</i>	32
4.2	Clonagem de genes de <i>C. jadinii</i> no vetor p416GPD.....	35
4.3	Transformação em <i>Saccharomyces cerevisiae</i> CEN.PK 113-7D	39
4.4	Caraterização fisiológica de <i>Cyberlindnera jadinii</i>	40
4.5	Expressão heteróloga do gene CT5 de <i>C. jadinii</i> em <i>S. cerevisiae</i>	44
4.6	Análise genómica comparativa.....	47
4.6.1	Ploidia.....	47
4.6.2	Anotação do genoma	48
5.	CONCLUSÃO E PERSPETIVAS FUTURAS.....	53
6.	BIBLIOGRAFIA	55
7.	ANEXOS.....	64

LISTA DE ABREVIATURAS E ACRÓNIMOS

YPD - *Yeast extract peptone dextrose medium*

LB - *Lysogeny Broth ou Luria-Bertani broth*

YNB - *Yeast Nitrogen Base*

DNA - Ácido desoxirribonucleico (do inglês *deoxyribonucleic acid*)

DNAg – DNA genómico

DNAp – DNA plasmídico

SDS - Dodecil-sulfato de sódio (do inglês *Sodium dodecyl sulfate*)

PCR – Reação em cadeia da polimerase (do inglês *polymerase chain reaction*)

BWA - *Burrows-Wheeler Alignment*

SNP - Polimorfismos de nucleótido único (do inglês *single nucleotide polymorphisms*)

NCBI - *National Center for Biotechnology Information*

Bps – Pares de bases

Mbp – Pares de megabases

v/v - Volume por volume

p/v – Peso por volume

g/L - Grama por litro

M - Molar

rpm – Rotações por minuto

OD_{640 nm} – Densidade ótica medida a um comprimento de onda de 640 nm

LISTA DE FIGURAS

Figura 1 - Exemplo modelo da linha de código utilizando a ferramenta <i>BCFtools filter</i> , com parâmetros definidos de exclusão, qualidade e cobertura do genoma.	30
Figura 2 - Exemplo da linha de código utilizando a ferramenta <i>BCFtools isec</i> para interseção dos ficheiros das quatro estirpes.	30
Figura 3 - Produtos de PCR da amplificação de genes de <i>C. jadinii</i> com a enzima Taq II.	34
Figura 4 - Produtos de PCR da amplificação de genes de <i>C. jadinii</i> com a enzima <i>Accuzyme</i>	35
Figura 5 - Digestão do plasmídeo p416GPD com a enzima <i>EcoRI</i>	36
Figura 6 - Clonagem dos genes CT5, CT6, CT7, CT10, CT11, CT12 e CT14 de <i>C. jadinii</i> no vetor p416GPD em células de <i>E. coli</i> XL-1 Blue.	37
Figura 7 - Transformação em células <i>S. cerevisiae</i> CEN.PK 113-7D do gene CT5 clonado no vetor p416GPD.	39
Figura 8 - Transformação em células <i>S. cerevisiae</i> CEN.PK 113-7D do gene CT6 clonado no vetor p416GPD.	40
Figura 9 - Padrão de crescimento de cinco estirpes de <i>C. jadinii</i> a 18 °C.	41
Figura 10 - Padrão de crescimento de cinco estirpes de <i>C. jadinii</i> a 30 °C.	42
Figura 11 - Fenótipos de crescimento da estirpe <i>S. cerevisiae</i> CEN.PK 113-7D transformada com o gene CT5 de <i>C. jadinii</i> clonado no vetor p416GPD.	46
Figura 12 - Determinação da ploidia de diferentes estirpes de <i>C. jadinii</i> utilizando o programa <i>nQuire</i>	48
Figura 13 - Variação nucleotídica de estirpes de <i>C. jadinii</i> contra o genoma de referência (<i>C. jadinii</i> CBS 1600).	50

LISTA DE TABELAS

Tabela 1- Alguns dos comandos do <i>SAMtools</i>	12
Tabela 2- Alguns dos comandos do <i>BCFtools</i>	13
Tabela 3- Identificação das estirpes de levedura utilizadas neste estudo.....	17
Tabela 4 - Meios de cultura suplementados com as diferentes fontes de carbono.	18
Tabela 5 - Componentes utilizados no processo de digestão de DNAp p416GPD com a enzima <i>EcoRI</i>	20
Tabela 6 - Identificação dos genes de <i>C. jadinii</i> em estudo neste trabalho.....	21
Tabela 7 - Oligonucleótidos utilizados para amplificação e clonagem dos genes de <i>C. jadinii</i>	22
Tabela 8 - Reações testadas para amplificação de DNAg com uso de diferentes enzimas.	23
Tabela 9 - Componentes utilizados no processo de digestão de DNAp p416GPD com a enzima <i>DpnI</i>	26
Tabela 10 - Genes de <i>C. jadinii</i> selecionados para estudo neste trabalho.	32
Tabela 11 - Variabilidade genética de estirpes de <i>C. jadinii</i> contra a referência <i>C. jadinii</i> CBS 1600	50
Tabela 12 - Correspondência genética de estirpes de <i>C. jadinii</i> contra o genoma de referência CBS 1600	51

1. INTRODUÇÃO

A biotecnologia assume um papel cada vez mais relevante no setor industrial. A utilização de enzimas, extratos celulares ou microrganismos em processos industriais é aplicada na produção de uma ampla gama de produtos, processos e tecnologias, que incluem desde biocombustíveis até alimentos geneticamente modificados e compostos químicos (Gavrilescu e Chisti 2005; Heux et al. 2015).

O principal objetivo da biotecnologia industrial é o desenvolvimento de aplicações de natureza industrial através de sistemas biológicos como alternativa aos métodos clássicos, envolvendo catalisadores químicos (Heux et al., 2015; Sousa-Silva et al., 2021). Idealmente, este processo deverá ser desenvolvido de uma forma mais eficiente e econômica, em comparação com os métodos convencionais, (Gavrilescu e Chisti 2005; Meyer 2011) assim como de uma forma mais sustentável. A biotecnologia industrial é frequentemente uma alternativa mais sustentável, devido não só ao processo de produção, mas também à monitorização e gestão de resíduos. A capacidade de utilizar matérias-primas renováveis, diminuir a pegada de carbono, minimizar a produção de resíduos e destruir com segurança poluentes para recuperação do ambiente, torna-a uma ferramenta importante para lidar com desafios ambientais e alcançar metas de sustentabilidade (Gong et al., 2017).

Contudo, é essencial avaliar diversas propriedades dos sistemas biológicos, especialmente dos microrganismos, para alcançar elevados desempenhos metabólicos nas fábricas celulares. Vários são os processos que contribuem para o aumento da eficiência destes processos, incluindo a otimização de fluxos metabólicos, a regulação da expressão de transportadores de membrana e o aumento da resistência a condições industriais desfavoráveis e compostos tóxicos (Sousa-Silva et al., 2021).

As fábricas celulares microbianas são consideradas mais relevantes entre as restantes pela sua versatilidade metabólica em comparação com organismos mais complexos, a sua eficiência e taxa de crescimento maior, facilidade de manipulação genética, entre outros (Gong et al., 2017). As leveduras são amplamente reconhecidas por desempenharem um papel fundamental nos processos de fermentação tradicionais, que incluem a produção de pão, cerveja e vinho. Atualmente, são também valorizadas como fontes alternativas de proteínas, enzimas e vitaminas com alto valor nutricional, além de apresentarem inúmeras aplicações na indústria alimentar e

farmacêutica. Presentemente estão descritas cerca de 2000 espécies de levedura (<https://theyeasts.org/>), desde a levedura modelo *Saccharomyces cerevisiae*, às menos conhecidas, mas igualmente relevantes, como a *Cyberlindnera jadinii*, levedura em estudo neste trabalho (Fonseca et al., 2008; Morata et al., 2018; Murali, D., & Dhandayuthapani, 2019; Santiago et al., 2021; Tahir et al., 2017).

1.1 A levedura *Cyberlindnera jadinii*

1.1.1 Fisiologia

As espécies de levedura pertencentes ao género *Cyberlindnera* possuem um teor proteico elevado, equivalente a cerca de metade do seu peso celular seco. Além disso, destacam-se pela diversidade de aminoácidos e abundância em vitaminas (como niacina e vitaminas do complexo B). Estas particularidades tornam a sua biomassa ideal como fonte de suplemento proteico tanto para alimentação animal como para o consumo humano (Martínez et al., 2018; Sousa-Silva et al., 2021). As suas características fermentativas permitem o seu crescimento numa diversidade de substratos derivados de biomassa, sendo capazes de assimilar glicose, arabinose, sacarose, entre outros (Sousa-Silva, 2021). *C. jadinii* é classificada como *Crabtree*-negativa pelo que alcança níveis mais elevados de rendimento celular em ambientes aeróbicos em comparação com espécies *Crabtree*-positivas. O efeito *Crabtree*-negativo promove o processo respiratório em detrimento da fermentação, possibilitando, desse modo, o desenvolvimento de fenótipos propícios à produção de proteínas de interesse biotecnológico (Sousa-Silva et al., 2021). Assim, esta levedura detém uma das maiores capacidades respiratórias entre as espécies de leveduras caracterizadas, tornando-a ideal para o crescimento em culturas contínuas (Bekatorou et al., 2006). *C. jadinii* apresenta, igualmente, um elevado grau de tolerância às alterações ambientais que ocorrem durante o processo de fermentação. Isso inclui resistência a temperaturas elevadas e capacidade de crescer numa ampla faixa de temperaturas, variando entre 19 e 37 °C. Demonstra ainda uma notável capacidade de sobrevivência em ambientes de pH ácido, sendo capaz de manter-se viável em valores de pH de ~3.5 durante longos períodos (Nie et al., 2005; Sousa-Silva et al., 2021; Wang et al., 2015). Estas características tornam esta levedura numa alternativa interessante às fábricas celulares microbianas já implementadas a nível industrial.

Outro aspeto importante é o facto desta levedura ser capaz de utilizar tanto hexoses como pentoses como substrato para produção de biomassa (SCP “*single-cell protein*”), bem como

múltiplos produtos químicos de base biológica, por exemplo ácidos orgânicos, aminoácidos e polipéptidos (Castrillo et al., 1996; Wang et al., 2015). Uma outra propriedade de destaque deste microrganismo é a capacidade para exportar proteínas para o meio extracelular (Sousa-Silva et al., 2022).

A levedura *C. jadinii* possui o estatuto de “GRAS” (do inglês *generally recognized as safe*) pela organização *The Food and Drug Administration* (FDA), por ser considerada segura e adequada para a produção de aditivos alimentares. Esta classificação baseia-se no seu histórico na indústria alimentar ao longo do tempo (Bekatorou et al., 2006; Boze, Moulin, e Galzy 1992; Buerth, Tielker, e Ernst 2016).

Assim, a relevância deste microrganismo na biotecnologia industrial resulta grande parte do seu potencial para produzir uma grande variedade de compostos, crescer em ambientes de baixo custo com elevada produtividade e demonstrar capacidade de assimilação de diversas fontes de carbono (Sousa-Silva et al., 2021), sendo que muitas podem ser provenientes de resíduos ou subprodutos industriais.

1.1.2 A evolução da nomenclatura de *C. jadinii*

A levedura *C. jadinii* teve várias denominações ao longo dos anos. Em 1932, era conhecida por *Saccharomyces jadinii*, em 1951 passou a ser denominada de *Hansenula jadinii*, seguida de *Candida utilis* em 1952. Por volta de 1984, o seu nome era *Pichia jadinii* e, em 2008 *Lindnera jadinii*, sendo esta a última designação antes do nome atualmente estabelecido. Desde 2009 é designada de *Cyberlindnera jadinii* (Buerth et al., 2016; Cletus P. Kurtzman et al., 2008; Rupp et al., 2015). No entanto durante anos persistiu a dúvida entre a relação genética existente entre *C. jadinii* e *C. utilis*. Vários cientistas realizaram uma análise abrangente das duas espécies ao nível genómico e fisiológico. Alguns aspetos que evidenciam a proximidade genética entre estas duas leveduras incluem uma taxa de reassociação de 85% entre o DNA genómico de ambas as espécies, notáveis semelhanças entre as sequências de RNA ribossómico e fenótipos de crescimento semelhantes (Buerth et al., 2016; Yamada et al., 1995). Algo que distingue as diferentes espécies testadas é o facto de *C. jadinii* esporular, quando cultivada em meio YPD, ao contrário de *C. utilis*, sugerindo assim que *C. jadinii* é a espécie parental teleomórfica de *C. utilis*, sendo esta a representação da forma de crescimento anamórfico assexuado (Buerth et al., 2016; C. P. Kurtzman et al., 1979; Rupp et al., 2015). Mais tarde, a relação genética entre as duas espécies foi finalmente confirmada pela

comparação das sequências genômicas completas da estirpe de *C. utilis* (ATCC9950) e de *C. jadinii* (CBS1600). Esse estudo demonstrou que o tamanho do genoma era semelhante, com aproximadamente 12.8 e 12.7 Mbp, respectivamente, e a identidade de sequência genômica era de 98% (Krahulec et al., 2020; Rupp et al., 2015). Atualmente, todas as diferentes estirpes são designadas por *C. jadinii*.

Relativamente à ploidia de *C. jadinii*, várias técnicas foram utilizadas para a sua determinação, tais como: observação da formação de ascósporos e análise de dados de sequenciação genômica (Sousa-Silva et al., 2021). Essa análise incluiu a detecção de polimorfismos de nucleotídeo único (SNP) e obtenção de 50% de correspondência quando realizado o alinhamento de múltiplas sequências de leitura com uma sequência de referência genômica, confirmando a diploidia da estirpe *C. jadinii* NRRL Y-1542 (Buerth et al., 2016; Krahulec et al., 2020; Sousa-Silva et al., 2021). No entanto, a análise de SNP evidenciou que, apesar do caráter predominantemente diploide do genoma, este pode conter regiões com diferentes níveis de ploidia, incluindo haploidia, triploidia e tetraploidia, resultante de eventos de duplicação e deleção de fragmentos. As regiões de maior ploidia encontradas no genoma de *C. jadinii* podem ser vantajosas, por exemplo, como locais de inserção e para melhorar a produção de proteínas heterólogas (Rupp et al., 2015).

A diploidia identificada em algumas estirpes de *C. jadinii* impulsionou a sua manipulação genética e posterior utilização no setor biotecnológico. Esta característica confere uma notável vantagem à manipulação genética desta estirpe perante a obtenção de mutantes, que são obtidos mais facilmente numa levedura diploide do que em leveduras com maiores estados de ploidia (Gu et al., 2023; Sousa-Silva et al., 2021).

1.1.3 Relevância biotecnológica

A expressão heteróloga de genes tem-se tornado cada vez mais importante para a produção de proteínas recombinantes, tanto na indústria farmacêutica como no setor biotecnológico.

A levedura *S. cerevisiae* é um dos hospedeiros de expressão mais utilizados, graças à sequenciação do seu genoma e à facilidade da sua manipulação genética. Porém, esta levedura, dita convencional, tem alguns inconvenientes como por exemplo, crescimento limitado em substratos com elevado teor de glucose e secreção pouco eficiente de proteínas. Por esse motivo, várias leveduras não convencionais foram estabelecidas como hospedeiras para a expressão de proteínas recombinantes, incluindo *Hansenula polymorpha*, *Pichia pastoris* e *Cyberlindnera jadinii*,

que foram utilizadas na produção de diversas proteínas de interesse farmacêutico ou biotecnológico (Buerth et al., 2016; Kunigo et al., 2013). Algumas leveduras, como é o caso de *C. jadinii*, são fábricas celulares eficientes que apresentam diversas vantagens, tais como segurança biológica, crescimento rápido, modificação genética simples, sistemas de secreção eficientes e modificações pós-tradução de proteínas (Buerth et al., 2016). *C. jadinii* possui uma capacidade metabólica que lhe permite utilizar diversas fontes de carbono como única fonte de carbono e energia (Sousa-silva, 2021).

O primeiro sistema de transformação da levedura *C. jadinii* foi desenvolvido em 1995 por Kondo et al. (Kondo et al., 1995). A sua utilização como hospedeiro para a expressão de proteínas heterólogas só foi desenvolvida nas últimas duas décadas, tendo sido adiada devido à falta de marcadores de seleção. Porém, mais recentemente foram desenvolvidos sistemas de transformação genética eficientes que permitiram a otimização da produção de produtos exógenos (Gu et al., 2023; Sousa-Silva et al., 2021).

A engenharia metabólica aplicada a *C. jadinii* revelou-se bem-sucedida e tem cada vez mais impacto para a produção de metabolitos de interesse comercial. Por exemplo, foram alcançados níveis elevados de lactato (95,1% de conversão da glicose) expressando a lactato desidrogenase bovina (*Bos taurus*) numa estirpe mutante de *C. jadinii* deletada na enzima piruvato descarboxilase (Ikushima, Fujii, Kobayashi, et al., 2009). Outros estudos desenvolveram estirpes capazes de produzir elevados níveis de metabolitos, incluindo etanol, isopropanol e carotenoides, utilizando a xilose como uma fonte de carbono (Buerth et al., 2016).

Algumas das proteínas recombinantes produzidas intracelularmente em *C. jadinii* incluem o adoçante monelina nesta levedura, a α -amilase e xilanase (Kondo et al., 1997). A produção de monelina foi superior a 50% da proteína solúvel, no caso da α -amilase bacteriana foram atingidos igualmente níveis elevados de produção, constituindo cerca de 27% das proteínas totais. Contrariamente, a produção de xilanase de *Streptomyces olivaceoviridis* XynA, originou níveis inferiores, quando comparado com outros sistemas de expressão (Buerth et al., 2016; Kondo et al., 1995) (Buerth et al., 2016; Kondo et al., 1997).

Esta levedura possui também capacidade de secretar proteínas heterólogas, pelo que secreta um conjunto limitado de proteínas para o meio de crescimento (Buerth et al., 2011). A secreção de proteínas recombinantes no meio é vantajosa, pois a sua purificação é facilitada pela menor quantidade de proteínas relativamente aos extratos celulares, além disso não é necessário lisar as

células) (Buerth et al., 2016; Kunigo et al., 2013). Um estudo de Kunigo et al., demonstrou que estirpes recombinantes de *C. jadinii* são capazes de secretar proteínas heterólogas no meio de crescimento, tais como a lipase B de *Candida antarctica* (CalB), xilanase do fungo *Penicillium simplicissimum* (XynA) e epítipo da insulina (Kunigo et al., 2013; Sousa-Silva et al., 2021). A secreção de xilanase XynA é relevante para o potencial de expressão de proteínas heterólogas de *C. jadinii*, pois permite o crescimento das células em xilano, como única fonte de carbono, substituindo assim a glucose e outras fontes de carbono mais dispendiosas (Kunigo et al., 2013, 2015). Esta capacidade aliada à tolerância às elevadas temperaturas e ambientes de pH ácidos, e ainda a eficácia da secreção de proteínas no meio de crescimento, tornam *C. jadinii* um hospedeiro promissor para aplicações biotecnológicas.

1.1.4 Aplicações industriais de *C. jadinii*

A aplicação desta levedura na indústria alimentar remonta a pelo menos sete décadas atrás (Bekatorou et al., 2006). Além de ser extensivamente empregue como aditivo ou substituto alimentar, *C. jadinii* desempenha um papel fundamental na produção de diversos suplementos, tais como vitaminas, proteínas, polissacarídeos e ácidos orgânicos. A sua utilização no fabrico de outros produtos, como maionese (Campos et al., 2015), bolo de azeite (Moftah et al., 2012) e bolachas (Ribeiro et al., 2020) etc., evidencia sua versatilidade e importância na indústria alimentar. Embora em pequenas quantidades, *C. jadinii* tem sido utilizada como aditivo na alimentação de vários animais pelas suas propriedades funcionais, assim como principal fonte de proteína na sua alimentação (Cruz, Sterten, et al., 2020; Cruz, Tauson, et al., 2020). Por exemplo, diferentes estirpes desta levedura foram avaliadas como substitutos a elementos específicos nas dietas para peixes, como é o caso do salmão do Atlântico (*Salmo salar*), evidenciando um elevado desempenho no seu crescimento (Cruz, Sterten, et al., 2020). E segundo estudos de Cruz et al. (2019), a inclusão de *C. jadinii* na dieta animal, representando entre 20 a 40% da proteína bruta, levou a um aumento da taxa de crescimento de alguns animais (Cruz et al., 2019, 2020; Lezcano e Herrera, 2013). Esta levedura apresenta ainda uma vasta gama de componentes bioativos, tais como β -glucanos e nucleótidos, que melhoram a função digestiva (Cruz et al., 2020; Kogan e Kocher, 2007).

Nos últimos anos, as aplicações deste microrganismo expandiram-se além do setor alimentar, abrangendo áreas como a cosmética e bem-estar, indústria química, agricultura e produção vinícola. Um exemplo é a capacidade de *C. jadinii* produzir o polissacarídeo β -D-glucano, o qual

pode ser utilizado no setor cosmético para formulação de diversos produtos, como cremes hidratantes, champô, produtos de limpeza facial e protetor solar. Isto deve-se às propriedades hidratantes do β -D-glucano, assim como resistência à radiação e oxidação, tornando-o um produto bio valioso neste contexto (Ma e Liu, 2018 a) b) c) d) e); Sousa-Silva et al., 2021).

Já no contexto agrícola, *C. jadinii* foi incorporada num fertilizante orgânico granular, resultando na otimização da eficiência da fertilização do solo, ao reduzir a quantidade necessária de fertilizante aplicado, proporcionando, simultaneamente uma libertação prolongada de nutrientes. Além disso, a aplicação desta levedura tem sido objeto de investigação para fins de biorremediação do solo, dada a sua eficiência na transformação de cádmio presente em solos contaminados, convertendo-o numa forma menos tóxica (Bai et al., 2017; Cui et al., 2015; Firgatovich et al., 2020).

Finalmente, na produção vinícola, esta levedura foi utilizada num processo de fermentação para criação de uma cerveja frutada, caracterizada pela ausência de teor alcoólico, assim como na produção de vinho de nêpera, com o intuito de reduzir a acidez e intensificar o aroma desta bebida (Bellut et al., 2019).

1.2 Proteínas transportadoras de membrana

Em qualquer processo bioquímico, a otimização do transporte de solutos através das membranas é muito importante para a produção de compostos biológicos com um elevado rendimento. Presentemente, a indústria de fermentação microbiana depara-se com dois grandes desafios na linha de produção. O primeiro diz respeito à acumulação intracelular e toxicidade do produto, juntamente com a produção insuficiente do mesmo no meio extracelular. O segundo desafio está relacionado com a capacidade intrínseca da fábrica celular para assimilar fontes de carbono e energia com a finalidade de promover a biossíntese do produto (Boyarskiy & Tullman-Ercek, 2015; Sauer et al., 2008). Estes dois obstáculos podem ser superados através da manipulação genética dos transportadores de membrana incluindo, a modulação da expressão de um importador e de um exportador, para otimizar o influxo de um substrato e o exporto do metabolito de interesse, respetivamente. Este processo poderá ainda beneficiar da regulação da atividade de um transportador endógeno ou exógeno, através do ajuste da sua especificidade, afinidade e/ou a capacidade do transportador. Essas intervenções poderão melhorar o consumo

do substrato e evitar a toxicidade do produto, levando a um melhor rendimento de produção (Boyarskiy & Tullman-Ercek, 2015; Kell et al., 2015; I. Soares-Silva et al., 2020).

Assim, as proteínas transportadoras desempenham um papel crucial na otimização de processos biotecnológicos, contribuindo para melhorar a entrada de nutrientes na célula, troca de produtos entre os diferentes organelos e o citoplasma, e o efluxo de metabolitos para o meio extracelular, aumentando a tolerância da célula a produtos finais tóxicos. Os transportadores podem também intervir na regulação do metabolismo e na resposta a estímulos ambientais (Casal et al., 2008; Kell et al., 2015; I. Soares-Silva et al., 2020).

Recentemente foram desenvolvidas novas estratégias para a identificação de transportadores com interesse biotecnológico. Estas incluem utilização de perfis de expressão, estudos de knockout em todo o genoma (onde a função dos genes é compreendida pela eliminação dos mesmos) e uma seleção baseada na resposta ao stress (Boyarskiy & Tullman-Ercek, 2015; Kell et al., 2015).

As leveduras são consideradas um dos grupos de microrganismos industriais mais promissores para o desenvolvimento de fábricas celulares, em particular para a produção de ácidos orgânicos, entre estas leveduras, destaca-se a *Cyberlindnera jadinii*. A sua versatilidade metabólica e capacidade de crescimento numa ampla gama de substratos torna-a um organismo de interesse para a biotecnologia, pelo que a compreensão dos transportadores de membrana nesta levedura é fundamental para perceber como ela utiliza nutrientes distintos em diferentes condições ambientais (Sousa-Silva, 2021; Sousa-Silva et al., 2021). Numa publicação de Soares-Silva et al. (2020) foram reportados sistemas de transporte membranar caracterizados em *C. jadinii* (importadores e exportadores). Entre alguns destes transportadores membranares constam o sistema de transporte do cobre (Cu^{2+}), do zinco (Zn^{2+}), do manganês (Mn^{2+}), do sulfato (SO_4^{2-}), e do nitrato (NO_3^-) (Ali & Hipkin, 1986; Benítez et al., 1983; Failla et al., 1976; Parkin & Ross, 1986). A absorção de todos estes elementos em *C. jadinii* é essencialmente dependente de energia e de pH (Sousa-Silva et al., 2021).

A caracterização fisiológica dos sistemas de transporte de ácidos orgânicos em levedura, levou à identificação de quatro sistemas de transporte mediado de ácidos orgânicos em *C. jadinii*, nomeadamente de um simporte de protões e monocarboxilatos, um segundo de dicarboxilatos, um de tricarboxilatos e por último, uma difusão facilitada de ácidos mono, di e tricarboxílicos. Todos estes transportadores de ácidos carboxílicos são induzidos pelos respetivos substratos, sendo sujeitos à repressão da glucose, bem como por concentrações de ácido superiores a 3%

(p/v) (Cássio & Leão, 1993). A difusão facilitada ocorre para a forma não dissociada dos ácidos carboxílicos, funcionando como uma permease geral de ácidos orgânicos (Cássio & Leão, 1993; ROUSH et al., 1959).

Recentemente, numa investigação de Sousa-Silva (2021), foram identificados novos transportadores de membrana desta levedura através da expressão heteróloga em *S. cerevisiae*. Esses transportadores pertencem a um total de seis famílias, cada uma com especificidade para ácidos orgânicos com relevância biotecnológica: *Acetate Uptake Transporter* (AceTr), *Sialate:H⁺ Symporter* (SHS), *Telurite-resistance/Dicarboxylate Transporter* (TDT), *Monocarboxylate Transporter* (MCT), *Solute:Sodium Symporter* (SSS) e *Divalent Anion:Na⁺ Symporter* (DASS). Estes transportadores apresentaram uma variabilidade significativa de especificidade, sendo capazes de transportar ácidos mono/ di e tricarboxílicos, tanto isoladamente quanto em combinação. Dezasseis deles transportam monocarboxílicos, cinco são responsáveis pelo transporte de dicarboxílicos (CjAto2p, CjAto5p, CjJen6p, CjSlc5p e CjSlc13-1p) e quatro são especializados no transporte de citratos (CjAto5p, CjJen6p, CjSlc5p e CjSlc13-1p) (Sousa-Silva, 2021).

Foram ainda analisados vários transportadores de aminoácidos do tipo simporte de prótons em *C. jadinii*, nomeadamente a arginina, lisina, glicina e glutamato. Assim como, trinta transportadores ABC de diferentes subfamílias de transportadores foram reconhecidos no genoma da estirpe NBRC0988 (estirpe homóloga à DSM2361) (Watanasrisin et al., 2016).

A expressão de transportadores de membrana aliada a ferramentas de engenharia metabólica é fundamental para desenvolver novas e mais eficientes estirpes para a produção de compostos de base biológica. Assim, um maior conhecimento das proteínas transportadoras permitirá o desenvolvimento de fábricas celulares otimizadas com impacto na biotecnologia industrial (Kell et al., 2015; Soares-Silva et al., 2020; Wang et al., 2016).

1.3 Estudo do genoma através da bioinformática

Nos últimos anos, com o constante progresso e evolução da biologia molecular, o aumento exponencial da informação biológica impôs a criação de um sistema mais estruturado e informatizado para recolher, armazenar, gerir e analisar a vasta quantidade de dados biológicos gerados em todas as áreas científicas (Tan et al., 2022; Zhang & Liu, 2013). O aumento da informação genómica levou a um crescimento paralelo da procura de ferramentas de armazenamento e gestão de dados avançadas e do desenvolvimento de *software* para análise,

modelização e previsão de grandes conjuntos de dados (Tan et al., 2022). Nesse sentido surgiu a bioinformática, como uma disciplina multidisciplinar que funde a biologia, a informática e a tecnologia numa única vertente. As suas aplicações vão desde sequenciação do genoma, a previsão da função e estrutura de proteínas, a filogenia, a conceção e o desenvolvimento de medicamentos e vacinas, até à compreensão da complexidade dos genes e do genoma (de Carvalho et al., 2019; Rao et al., 2008).

A aplicação de algoritmos de bioinformática na pesquisa e análise de sequências, combinada com outras técnicas, constitui uma abordagem rápida para obter informações detalhadas sobre genes e proteínas que são dificilmente obtidas por outras técnicas (Kaikabo & Kalshingi, 2008). A contribuição dos avanços da bioinformática tornou possível o mapeamento de todo o genoma humano e dos genomas de muitos outros organismos em pouco mais de uma década (Rao et al., 2008). Assim, a investigação bioinformática demonstra utilidade também no sentido de redução de tempo e custos associados.

O *The National Center for Biotechnology Information* (NCBI) mantém uma crescente coleção de bases de dados que reúnem informações sobre a sequência genética e a estrutura/atividade das proteínas, entre outros conjuntos de dados biológicos. Esta coleção progride a um ritmo exponencial (Rao et al., 2008).

Deste modo, a bioinformática é uma ferramenta essencial para desvendar os segredos codificados no genoma e o seu potencial para impulsionar descobertas científicas e avanços médicos é inestimável. O seu avanço não só é favorável para o domínio da biotecnologia, como também contribuirá largamente para o futuro da humanidade (Tan et al., 2022).

1.3.1 Softwares bioinformáticos

Para concretizar uma análise genómica abrangente intervêm aspetos fundamentais, como a anotação do genoma, previsão de regiões codificantes e determinação da ploidia. A comparação de vários genomas relacionados ou contra um genoma de referência pode ajudar a melhorar a sua anotação. Ao longo dos anos vários *softwares* têm surgido para a anotação e análise de genomas, nas suas diferentes vertentes. *Burrows-Wheeler Alignment* (BWA) é um *software* de alinhamento de sequências de DNA desenvolvido para mapear principalmente sequências curtas (*short reads*) e pouco divergentes em relação a um genoma de referência (Li & Durbin, 2009). É amplamente utilizado em análises de dados de sequenciação de próxima geração (NGS) e desempenha um

papel crucial em investigação em genómica e variação genética. Consiste em três algoritmos, o BWA-*backtrack* (Li & Durbin, 2009), BWA-SW (Li & Durbin, 2010) e BWA-MEM (Li, 2013). O primeiro apropriado para “reads” até 100 bp e os restantes para sequências maiores, entre 70 bp a 1 Mbp. O BWA-MEM (o mais recente) é geralmente o mais recomendado pela sua performance rápida e precisa. Quando comparado a outras ferramentas semelhantes, o algoritmo BWA-MEM demonstrou melhor desempenho (Abuín et al., 2015; Sarmiento et al., 2022). O *software* BWA suporta o alinhamento de sequências de uma única extremidade (*single-end reads*) com falhas, o que é cada vez mais importante quando se trata de mapear grandes volumes de *reads* e sequências que tendem a conter *indels* (inserções ou deleções). Além disso, esta ferramenta de alinhamento revelou-se cerca de 10 a 20x mais rápida que outras ferramentas com a mesma função, como por exemplo a MAQ (Li, 2012; Li & Durbin, 2009). O *output* desta ferramenta consiste em ficheiros no formato padrão SAM (*Sequence Alignment/Map*). A identificação de variantes e outras análises subsequentes após o alinhamento com BWA podem ser efetuadas com o *software* SAMtools. Estes dois *softwares* contêm a maioria das funcionalidades com características adicionais em comparação com outras ferramentas (Li et al., 2009; Li & Durbin, 2009). Desta forma devem ser utilizados de forma sequencial.

O SAMtools é um *software* amplamente utilizado, que consiste num conjunto de programas e ferramentas destinados ao processamento e análise de dados de sequenciação de elevado rendimento. Inclui ferramentas para consulta, ordenação, conversão e manipulação de ficheiros no formato SAM (*Sequence Alignment/Map*) e BAM (*Binary Alignment/Map*). Além disso, tem capacidade de remover duplicados originados em ensaios *PCR*, identificar *SNP's* e *indels* (Li et al., 2009; Ramirez-Gonzalez et al., 2012; Wright et al., 2019). Esta conversão entre tipos de ficheiros traz vantagens ao nível de uma maior facilidade de utilização, capacidade de ser suficientemente flexível para armazenar toda a informação de várias plataformas de sequenciação, tamanho compactado e capacidade de executar a maioria das operações sem a necessidade de carregar todo o conjunto de dados de alinhamento (Etherington et al., 2015; Li et al., 2009). Este *software* dispõe de uma variedade de comandos específicos (Tabela 1) para manipular dados, entre os quais se destacam o *view*, o *sort*, o *merge*, e *split*. O primeiro efetua a conversão de formatos, a filtragem de ficheiros e a extração de intervalos de sequência. Os seguintes são utilizados para ordenação, junção e divisão de ficheiros, respetivamente. É possível ainda aceder a qualquer ficheiro através desta ferramenta desde que seja indexado previamente, através do comando “*index*” (Danecek et al., 2021; Li et al., 2009).

A tabela que se segue evidencia alguns dos comandos existentes, as suas datas de criação e descrições respetivas.

Tabela 1- Alguns dos comandos do *SAMtools*. (Tabela adaptada de Danecek et al., (2021)).

Comando	Data	Descrição
<i>view</i>	22/12/2008	Conversão de ficheiros SAM/BAM
<i>sort</i>	22/12/2008	Ordenação de ficheiros
<i>merge</i>	22/12/2008	Junção de ficheiros ordenados
<i>index</i>	22/12/2008	Indexar ficheiros
<i>mpileup</i>	12/06/2010	Processamento de dados através de “ <i>pileup’s</i> ”
<i>split</i>	10/03/2014	Divisão de ficheiros

Para determinar a ploidia de diferentes estirpes de *C. jadinii* poder-se-iam utilizar diferentes *softwares* que incluem o *ConPADE* (Margarido & Heckerman, 2015), o *ploidyNGS* (Ferreira et al., 2019), e o *nQuire* (Weib et al., 2018), porém este último é dos mais eficientes e utilizados. Este *software* estatístico prevê diploidia, triploidia e tetraploidia ao realizar análise genómica comparativa. Embora a ploidia seja tradicionalmente determinada pela medição do conteúdo de DNA nuclear através do uso de citometria de fluxo (Dirihan et al., 2013; Qiu et al., 2020), é igualmente possível inferi-la de forma indireta (sem medição do conteúdo de DNA), a partir de dados de sequenciação de próxima geração (NGS). Esta distinção pode ser observada seja examinando distribuições k-mer (substrings de comprimento k contidos numa sequência biológica, constituídos por nucleótidos) ou avaliando a distribuição de frequências alélicas em polimorfismos de nucleótidos simples (SNPs) bialélicos (Weib et al., 2018). Esta metodologia tem sido utilizada para estimar a ploidia em genomas recentemente agrupados, a fim de identificar o número de haplótipos, bem como para detetar variações intraespecíficas de ploidia (Margarido & Heckerman, 2015). Tem também a vantagem de identificar aneuploidias (Libkind et al., 2021). Esta ferramenta modela a distribuição de frequências de base em locais variáveis usando um Modelo de Mistura Gaussiano (*Gaussian Mixture Model*), e através da máxima probabilidade seleciona o modelo de ploidia mais plausível. *nQuire* lida eficientemente com genomas extensos com alta cobertura e utiliza ficheiros com um *input* padrão (*short-reads* mapeadas) (Čadež et al., 2021;

Delomas et al., 2020; Weib et al., 2018). Uma desvantagem deste *software* é o facto de necessitar de uma genoma de referência para realizar as análises.

Para avaliar variabilidade entre as estirpes testadas foi usado o *software BCFtools*. Este programa combina uma série de utilidades que promovem o processamento de dados e deteção de variantes genéticas, como SNPs e *indels*. Inclui ferramentas de *variant calling* e de manipulação de ficheiros no formato VCF (*Variant Call Format*) e a sua contraparte BCF. Pode ser utilizado para indexar e ordenar ficheiros, adicionar ou remover anotações, comparar vários ficheiros simultaneamente com parâmetros personalizados, entre outros (Danecek et al., 2011, 2021; Narasimhan et al., 2016). Atualmente, o *BCFtools* é um programa completo que consiste em 21 comandos (Tabela 2) e 38 *plugins* (ferramentas de uma única finalidade), com mais de 230 opções e linhas de comando documentadas (Danecek et al., 2021). Todos os comandos aceitam os formatos VCF e BCF, tanto de uma forma comprimida como não comprimida. Ficheiros que não foram indexados geralmente também são lidos, mas não em todas as situações, portanto sempre que vários ficheiros VCFs são lidos simultaneamente, devem ser indexados e, também, compactados (Danecek et al., 2011).

Ocasionalmente, o programa *SAMtools* pode apresentar um desempenho mais lento durante a análise e pode ser ineficiente para o armazenamento de dados. No entanto, o *BCFtools* proporciona um formato mais eficiente para o processamento de dados (Danecek et al., 2021).

Tabela 2- Alguns dos comandos do *BCFtools*. (Tabela adaptada de Danecek et al., (2021)).

Comando	Data	Descrição
<i>view</i>	17/05/2012	Conversão VCF/BCF, visualização e filtragem
<i>merge</i>	07/08/2012	Junção de ficheiros VCF/BCF que não se sobrepõem
<i>isec</i>	06/09/2012	Interseções de ficheiros VCF/BCF
<i>filter</i>	12/02/2013	Filtrar ficheiros VCF/BCF através de limites fixos
<i>norm</i>	12/03/2013	Normalização de <i>indels</i> , alinhamento à esquerda
<i>call</i>	20/08/2013	Modelos de <i>variant calling</i> -m e -c
<i>index</i>	07/02/2014	Indexar ficheiros VCF/BCF
<i>mpileup</i>	05/08/2016	Transferido do programa <i>SAMtools</i>

Snippy é um *software* bioinformático especializado em *variant calling* para dados de sequenciação NGS. Tem como principal finalidade a identificação de mutações, como substituições (SNPs) e inserções e/ou deleções (*indels*), entre genomas haploides e um genoma de referência. Este *software* utiliza uma ferramenta para alinhar as *reads* com o genoma de referência e uma abordagem distinta (“*call*”) para averiguar se alguma das discrepâncias resultantes entre os genomas são reais ou aspetos que podem ser ignorados. Por fim, o *Snippy* analisa o efeito que cada diferença provoca nos genes previstos (como truncamento, mudança de estrutura) com outra ferramenta distinta, a fim de perceber se pode potencialmente afetar o fenótipo (Seemann, 2015; Yoshimura et al., 2019). O seu processo de análise é rápido, realizado num curto período e executado de forma eficiente, o que é particularmente importante em análises genómicas e bioinformáticas de grandes conjuntos de dados. A identificação de variantes em genomas haploides é relativamente mais simples em comparação com genomas diploides, o que pode permitir uma análise mais rápida e direta (Seemann, 2015; Yoshimura et al., 2019). Esta ferramenta é relativamente mais recente, sendo a sua primeira versão pública lançada em 2015 por Torsten Seemann (Seemann, 2015). No geral, requer como *input* um genoma de referência no formato FASTA ou GENBANK e os resultados são apresentados de variadas formas conforme a seleção de *output* (tabelas comparativas, formato BAM/VCF, ficheiro de registo com comandos executados, etc). Possui ainda a opção de definir parâmetros avançados para a análise dos genomas, tais como “*mapping quality*”, “*minimum coverage*”, “*minumum proportion for variant evidence*”, entre outros (Hiltemann et al., 2023; Syme, A., Gladman, S., & Seemann, T., 2023).

O *Snippy* possui a mais-valia de poder ser complementado com outras ferramentas de análise genómica, como o *Snippy-core*. Esta ferramenta requer a aplicação simultânea de vários ficheiros resultantes do *Snippy*, utilizando a mesma referência, para criar um alinhamento central. O *Snippy-core* analisa o nucleótido presente em cada local, em todas as amostras e no genoma de referência, que pode ser o mesmo (monomórfico) ou diferente em algumas estirpes (polimórfico ou variante). Ao juntar todos os locais de variação criam o “*core SNP genome*” (Hiltemann et al., 2023; Syme et al., 2023). Este processo é essencial no sentido de eliminar as variações genéticas não informativas e produzir um alinhamento mais compacto e focado nas regiões conservadas, auxiliando a investigação sobre relações evolutivas entre diferentes estirpes.

Todos estes *softwares* bioinformáticos têm sido amplamente utilizados para processar e analisar dados de sequenciação de todo o tipo de espécies, desde vertebrados, invertebrados,

patogénicos, plantas e vírus, o que proporciona desafios e oportunidades para futuros desenvolvimentos (Danecek et al., 2021).

Assim, a prática das abordagens bioinformáticas descritas é fundamental para compreender as características genómicas de microrganismos, como *C. jadinii*. Através de uma análise *in silico* que inclua estes métodos é possível determinar com eficácia a ploidia, realizar a anotação do genoma e detetar diferenças entre variadas estirpes. É possível ainda perceber relações existentes entre o comportamento fenotípico de diferentes estirpes e o seu genoma.

2. OBJETIVOS

A levedura *Cyberlindnera jadinii* desempenha um papel cada vez mais importante no setor biotecnológico. A sua elevada versatilidade, assim como a capacidade de utilização de uma ampla gama de ácidos orgânicos como única fonte de carbono e energia torna-a uma escolha atrativa para a produção de diversos compostos de interesse comercial. Estas características aliadas à sua eficiência na produção de proteínas recombinantes tornam esta levedura num microrganismo de interesse com elevado potencial biotecnológico. Assim, o objetivo principal deste trabalho foi estudar as características fisiológicas e genómicas de *C. jadinii* de modo a aprofundar o conhecimento sobre esta levedura, assim como potenciar a utilização desta levedura como fábrica celular microbiana.

Ao longo do trabalho foi realizada a caracterização fisiológica de diferentes estirpes de *C. jadinii*, tendo em vista a sua capacidade de utilização de diferentes ácidos carboxílicos. Potenciais transportadores de ácidos carboxílicos, identificados previamente numa análise do transportoma de *C. jadinii*, foram estudados neste trabalho, por expressão heteróloga em *Saccharomyces cerevisiae*.

Foi ainda realizada uma análise genómica comparativa de cinco estirpes de *C. jadinii*, nomeadamente PYCC 3092, PYCC 2578, TB104 e TB115, com anotação do seu genoma e determinação da ploidia através da utilização de ferramentas bioinformáticas. Adicionalmente, identificou-se a presença de SNP's e indels. O genoma da estirpe *C. jadinii* CBS 1600 foi utilizado como referência nos procedimentos executados.

Com este trabalho pretende-se aumentar o conhecimento genético, metabólico e fisiológico sobre este microrganismo, que poderá fortalecer o papel de *C. jadinii* no setor da biotecnologia.

3. MATERIAIS E MÉTODOS

3.1 Material biológico

Neste trabalho foram utilizadas diferentes estirpes de *Cyberlindnera jadinii*. Como estirpes de referência foram usadas as estirpes DSM 2361 e CBS 1600, e para análise comparativa dos genomas utilizaram-se as estirpes *C. jadinii* PYCC 3092, PYCC 2578, TB104 e TB115 (Tabela 3). Estas duas últimas foram isoladas no âmbito do projeto TransBio (Grant agreement ID: 289603) e encontram-se depositadas na coleção microbiológica do Departamento de Biologia da Universidade do Minho. Adicionalmente, foi também utilizada a levedura *Saccharomyces cerevisiae* CEN.PK 113-7D, para expressão heteróloga de transportadores de carboxilatos putativos de *C. jadinii*. A estirpe de *Escherichia coli* D4-SX foi utilizada para extração do plasmídeo p416GPD e a estirpe *E. coli* XL-1 Blue para as transformações. Todos os materiais biológicos foram armazenados em glicerol 30% (p/v) a uma temperatura de -80 °C.

O plasmídeo p416GPD foi utilizado como vetor de clonagem para genes de *C. jadinii* e vetor de expressão de *S. cerevisiae*.

Tabela 3- Identificação das estirpes de levedura utilizadas neste estudo.

Identificação	Espécie	Código da coleção	Referencia
I	<i>Cyberlindnera jadinii</i>	CBS 1600	Sousa-Silva, 2021
II	<i>Cyberlindnera jadinii</i>	DSM 2361	Sousa-Silva, 2021
III	<i>Cyberlindnera jadinii</i>	PYCC 2578	Sousa-Silva, 2021
IV	<i>Cyberlindnera jadinii</i>	PYCC 3092	Sousa-Silva, 2021
V	<i>Cyberlindnera jadinii</i>	TB104	Sousa-Silva, 2021
VI	<i>Cyberlindnera jadinii</i>	TB115	Sousa-Silva, 2021
VII	<i>Saccharomyces cerevisiae</i>	CEN.PK 113-7D	(Entian & Kötter, 2007)

3.2 Meios de cultura

Neste trabalho utilizaram-se diferentes meios de cultura para o crescimento dos microrganismos, mencionados de seguida:

- YP (extrato de levedura 0,5% p/v, peptona 1,0% p/v, agar 2,0% p/v) para o crescimento de culturas de leveduras. Este meio foi suplementado com diferentes fontes de carbono;
- YNB (YNB 0,67 g/100 mL, agar 2,0% p/v) meio mínimo para o crescimento de transformantes. Este meio foi suplementado com diferentes fontes de carbono;
- LB (extrato de levedura 0,5% p/v, triptona 1,0 % p/v, cloreto de sódio 0,5% p/v e agar 2,0 % p/v) para o crescimento de culturas bacterianas.

As culturas de leveduras foram mantidas em meio YPD (extrato de levedura 0,5% p/v, peptona 1% p/v, glicose 0,5% p/v, agar 2% p/v). Para os ensaios de crescimento das diferentes estirpes de *C. jadinii* utilizou-se o meio YP, ao qual foram adicionados diferentes ácidos carboxílicos como fonte de carbono e energia, nomeadamente, ácido cítrico, ácido láctico, ácido pirúvico, ácido fumárico, ácido málico, ácido acético, ácido succínico, e como controlo glucose (Tabela 4). Foi igualmente utilizado meio YNB, para avaliação de características fisiológicas dos transformantes. A tabela que se segue apresenta os meios elaborados.

Tabela 4 - Meios de cultura suplementados com as diferentes fontes de carbono.

Meio YNB	Meio YP
YNB + Ácido Cítrico (1,0% p/v), pH 6.0	YP + Ácido Cítrico (1,0% p/v), pH 6.0
YNB + Ácido Láctico (0,5% v/v), pH 5.0	YP + Ácido Láctico (0,5% v/v), pH 5.0
YNB + Ácido Pirúvico (0,5% p/v), pH 6.0	YP + Ácido Pirúvico (0,5% p/v), pH 6.0
YNB + Ácido Fumárico (1,0% p/v), pH 6.0	YP + Ácido Fumárico (1,0% p/v), pH 6.0
YNB + Ácido Málico (1,0% p/v), pH 6.0	YP + Ácido Málico (1,0% p/v), pH 6.0
YNB + Ácido Acético (0,5% v/v), pH 6.0	YP + Ácido Acético (0,5% v/v), pH 6.0
YNB + Ácido Succínico (1,0% p/v), pH 6.0	YP + Ácido Succínico (1,0% p/v), pH 6.0
YNB + Glucose (2% p/v)	YP + Glucose (2% p/v)

O pH dos meios foi ajustado com uma solução de KOH 10 M. Os meios de cultura foram submetidos a esterilização em autoclave a 120 °C, durante 20 minutos, a 1 atm.

3.3 Condições de crescimento

Para extração de DNA genómico a estirpe *C. jadinii* DSM 2361 foi inoculada em meio sólido YPD e incubada a 30 °C durante 24 h. Uma pequena quantidade de biomassa foi transferida para um tubo cónico de 50 mL contendo 10 mL de meio YPD, que foi posteriormente incubado a 30 °C, com agitação a 120 rpm, durante a noite.

A estirpe *E. coli* D4-SX foi inoculada em placa com meio LB contendo ampicilina incubada de seguida na estufa a 37 °C por 24 h. O inóculo em meio líquido foi efetuado num tubo cónico de 50 mL contendo 10 mL de meio LB com ampicilina (1% v/v) ao qual se acrescentou uma porção de biomassa das células de *E. coli*. A cultura foi incubada a 37 °C, 200 rpm, durante a noite.

Todos estes processos foram executados em condições de assepsia.

3.4 Preparação de células competentes de *Escherichia coli*

As células competentes da estirpe *E. coli* XL-1 Blue, foram preparadas utilizando o protocolo *Rubidium Chloride* (RbCl₂) de Sambrook, Fritsch e Maniatis (1989). Adicionalmente, foram utilizadas células competentes da NZYTEC (Portugal).

3.5 Extração de DNA genómico de *Cyberlindnera jadinii*

Cerca de 4 colónias de *C. jadinii* DSM 2361, foram suspensas em 100 µL de 0,2 M LiOAc 1% SDS e incubadas a 70 °C durante 15 min. Seguidamente foram adicionados 300 µL de etanol 96% (v/v) ao microtubo, sendo esta solução misturada vigorosamente no agitador de tubos, e centrifugada de seguida por 3 min a 15000 *g*, descartando por fim o sobrenadante. Adicionou-se 300 µL de etanol 70% (v/v) e centrifugou-se nas mesmas condições descritas anteriormente. Descartou-se o sobrenadante e resuspendeu-se o sedimento em 100 µL de H₂O ultrapura. Por fim, centrifugou-se a suspensão durante 3 min à velocidade máxima. O sobrenadante foi transferido para um novo microtubo, uma vez que é no sobrenadante onde se encontra o DNA genómico. O DNAg foi quantificado com o espectrofotómetro NanoDrop ND-1000 (Thermo Scientific, USA) e seguidamente foi armazenado a -20 °C (Lööke et al., 2011).

3.6 Extração de DNA plasmídico

O DNA plasmídico foi extraído de *E. coli* com o Kit NZYMiniprep (NZYTech, Portugal). Foram usados cerca de 5 mL do inóculo realizado previamente (secção 3.3) e foram seguidas as instruções do fabricante. Por fim, a amostra de DNA plasmídico foi quantificada utilizando o NanoDrop ND-1000 (Thermo Scientific, USA) e guardada a -20 °C.

3.7 Digestão de DNA plasmídico com enzimas de restrição

Após a extração do DNAp p416GPD realizou-se a digestão do mesmo com enzimas de restrição, para posteriormente realizar recombinação genética, inserindo o gene de interesse no plasmídeo linearizado (secção 3.10). Inicialmente foi usada a enzima *EcoRI* (ThermoFisher Scientific, USA). O protocolo de digestão e os respetivos componentes estão mencionados na Tabela 5, abaixo:

Tabela 5 - Componentes utilizados no processo de digestão de DNAp p416GPD com a enzima *EcoRI*.

Componente	Quantidade (µL)
Enzima <i>EcoRI</i> (10 U/µL)	1
Tampão (tango 2x concentrado)	4
Amostra de DNAp p416GPD ^a	1
H ₂ O U.P.	13
Fosfatase Alcalina (1 U/µL)	1
Total	20

^a – Concentração DNAp = 779,3 ng/µL

Foi adicionada a enzima fosfatase alcalina (NZYTech, Portugal) a esta reação para impedir a religação dos fragmentos de DNA digerido. Depois a reação foi incubada a 37 °C durante 16h. O protocolo em questão foi baseado na informação disponível no ThermoFisher Scientific (<https://www.thermofisher.com/order/catalog/product/ER0271>).

Posteriormente foram testadas as enzimas KpnI, a XbaI, a SmaI, HindIII (ThermoFisher Scientific, USA), com o intuito de verificar se possuíam maior eficácia na digestão do plasmídeo p416GPD, e evitar a religação do mesmo.

3.8 Purificação de DNA plasmídico de um gel de agarose

A purificação do DNA p416GPD linearizado foi efetuada com o Kit Gelpure (NZYTech, Portugal), de acordo com as instruções do fabricante. Para isso 60 µL de amostra de DNA linearizado, foi transferida para um tubo de 1.5 mL e adicionou-se 5x o volume desta de tampão de ligação (*Binding Buffer*). Neste caso foram utilizados 300 µL de tampão. Por fim, a amostra de plasmídeo digerido e purificado foi quantificada utilizando, mais uma vez, o NanoDrop ND-1000 (Thermo Scientific, USA), e guardada a -20 °C. A amostra tinha uma concentração de 25,8 ng/µL.

3.9 Amplificação do DNA e análise por eletroforese em gel de agarose

Neste trabalho foi estudado um conjunto de 8 genes da levedura *C. jadinii* (Tabela 6), os quais foram amplificados para posterior clonagem no vetor p416GPD. Para se proceder à amplificação preparou-se, previamente, os oligonucleótidos a utilizar (Tabela 7).

Tabela 6 - Identificação dos genes de *C. jadinii* em estudo neste trabalho.

Gene	Número de acesso
CT5	XP_020067736.1
CT6	XP_020068615.1
CT7	XP_020068161.1
CT10	XP_020070234.1
CT11	XP_020070372.1
CT12	XP_020068637.1
CT13	XP_020069005.1
CT14	XP_020068827.1

Tabela 7 - Oligonucleótidos utilizados para amplificação e clonagem dos genes de *C. jadinii*.

Gene	Primer	Sequência	Tm
CT5	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGAATGAATAAAGAAGTCGTGTC	50.94
CT5	Rev	ACTAATTACATGACTCGAGGTCGACGGTATTCAAATCTTCACATTGAGCA	53.10
CT6	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGAATGAGTAAAGTCAAGAACTT	49.52
CT6	Rev	ACTAATTACATGACTCGAGGTCGACGGTATTCACATCCTGTAGTTAAAAA	49.51
CT7	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGAATGTCTTTGAACGATAAATC	49.14
CT7	Rev	ACTAATTACATGACTCGAGGTCGACGGTATTCAGCTGGCCACGCCTCGCT	69.16
CT10	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGAATGGATAGATTTGTCACTCT	50.89
CT10	Rev	ACTAATTACATGACTCGAGGTCGACGGTATTTAGAGGGTGAAGACGCTTG	57.26
CT11	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGACATCGCACTGTTAATATCAC	52.22
CT11	Rev	ACTAATTACATGACTCGAGGTCGACGGTATCTAGAACAACCACGACAGGA	56.90
CT12	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGAATGACTGATCTCGAGTCCGG	58.69
CT12	Rev	ACTAATTACATGACTCGAGGTCGACGGTATTTAGTTCTTCTCAGCAGATT	51.50
CT13	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGAATGTCGACAGGTCAAAGTGA	56.80
CT13	Rev	ACTAATTACATGACTCGAGGTCGACGGTATTCACCTTCTTCTTCTTGA	51.36
CT14	Fwd	CACCAGAACTTAGTTTCGACGGATTCTAGAATGGTCAAGTCAACCTTGTA	54.18
CT14	Rev	ACTAATTACATGACTCGAGGTCGACGGTATTCATTCTAACTTATCGTAGA	47.96

Os genes indicados na tabela anterior foram amplificados por PCR (*Polymerase Chain Reaction*), utilizando o DNA genómico da levedura *C. jadinii* DSM 2361. Inicialmente, foi realizada uma reação utilizando a enzima Taq II DNA polimerase (NZYTech, Portugal), de acordo com a Tabela 8. Posteriormente, efetuou-se a amplificação dos 8 genes por PCR utilizando a enzima *Accuzyme* DNA polimerase, com atividade *proofreading* (Bioline, London UK) (Tabela 8). Este procedimento foi realizado separadamente para cada gene, utilizando os *primers* respetivos listados na Tabela 7. A tabela seguinte apresenta as diferentes reações testadas, incluindo os reagentes utilizados em cada (para informações detalhadas consultar os anexos – Tabela A1).

Tabela 8 - Reações testadas para amplificação de DNAg com uso de diferentes enzimas.

Reação 1	Reação 2
Taq II DNA polimerase	Accuzyme (2x concentrada)
Tampão (10x concentrado)	-
MgCl ₂ (50 mM)	-
dNTPs (100 mM)	-
DNAg	DNAg
H ₂ O U.P.	H ₂ O U.P.
<i>Primer forward</i> (10 µM)	<i>Primer forward</i> (10 µM)
<i>Primer reverse</i> (10 µM)	<i>Primer reverse</i> (10 µM)

Os vários ensaios de PCR foram elaborados sob as seguintes condições: uma fase inicial de desnaturação a 95 °C por 5 minutos, seguido de 35 ciclos constituídos por: desnaturação (95 °C por 30 segundos), hibridização (variável) e extensão (72 °C por 1 minuto). Cada ciclo foi concluído com uma extensão final de 5 minutos a 72 °C. De referir que, no caso da reação utilizando a enzima *Accuzyme*, o tempo definido para a extensão a 72 °C dentro dos 35 ciclos foi estendido para 3 minutos, em vez de 1.

As temperaturas de hibridização foram escolhidas com base nas temperaturas de desnaturação (*T_m* - *melting temperature* em inglês) de cada *primer* (conforme indicado na tabela 7). Portanto, durante essa etapa do processo de PCR utilizaram-se, inicialmente, gradientes de temperatura para ajustar as condições ideais a cada gene, otimizando posteriormente de acordo com os resultados.

Para a análise dos resultados, os produtos de PCR foram submetidos ao processo de eletroforese em gel de agarose (1%). Na maioria dos ensaios, 15 µL de amostra de cada gene amplificado foram adicionados a 3 µL de *loading buffer* (6x) e a totalidade desta solução foi aplicada no gel. Os géis correram por um período de 35 a 40 minutos a 90 Volts, utilizando o marcador de DNA, "*Green Safe*" (NZYTech, Portugal). Para a visualização das bandas no gel foi utilizado o ChemiDoc (Biorad, Italy). Por fim, os resultados foram quantificados com o NanoDrop ND-1000 (Thermo Scientific, USA).

3.10 Clonagem dos genes de *C. jadinii* no vetor p416GPD

Uma vez alcançada a amplificação pretendida dos produtos PCR referentes aos genes em estudo, procedeu-se à etapa de clonagem no vetor de expressão p416GPD. Para o efeito utilizaram-se células competentes de *E. coli* XL-1 Blue (NZYTech, Portugal), adotando o método *Gap Repair* proposto por Jacobus e Gross (2015).

Com base neste método, foi elaborada uma mistura com 200 *ng* do fragmento de inserção e 100 *ng* de vetor, para cada gene. Esta mistura foi adicionada aos tubos contendo as células competentes previamente descongeladas em gelo. Posteriormente, esta solução foi submetida a um período de incubação de 30 minutos no gelo. Consequentemente, realizou-se um choque térmico, no qual as amostras foram transferidas do gelo para um termobloco a 42°C por um período de 1 minuto, agitando os tubos nos instantes finais. As amostras foram recolocadas no gelo por mais 10 minutos. Para viabilizar o crescimento das células, 800 μ L de meio de cultura LB, previamente aquecido, foi adicionado a cada tubo de células, as quais foram incubadas a 37 °C e 200 rpm/minuto, durante 1 hora. Após esse período, procedeu-se à centrifugação das células durante 1 minuto a 14 000 g, seguido de descarte de 200 μ L de sobrenadante. Com as células devidamente ressuspensas e homogeneizadas, a solução resultante foi distribuída e inoculada em placas contendo meio de cultura LB suplementado com ampicilina (1x concentrada) como antibiótico seletivo. As referidas placas foram então incubadas a 37 °C, durante a noite, visando promover o desenvolvimento e crescimento das células transformadas em *E. coli*.

Foi, simultaneamente, efetuado um controlo negativo apenas com DNA plasmídico p416GPD.

3.10.1 Confirmação da clonagem no plasmídeo p416GPD

Com o objetivo de identificar possíveis clones positivos, foi executada a extração de DNA plasmídico dos transformantes de *E. coli* obtidos, utilizando o Kit NZYMiniprep mencionado na secção 3.5. As amostras de DNA plasmídico extraído foram, posteriormente, submetidas a um PCR com 35 ciclos utilizando a enzima Taq II DNA polimerase (conforme descrito na secção 3.9) e o par de *primers* (Tabela 7) do gene correspondente, a fim de amplificar os fragmentos de inserção e confirmar se os mesmos tinham o tamanho pretendido. Para a visualização dos fragmentos obtidos foi empregue a técnica de eletroforese em gel de agarose.

3.10.2 PCR de colónia

Neste ensaio efetuou-se a técnica PCR, utilizando a biomassa das colónias transformantes de *E. coli* como fonte de DNA, com o propósito de verificar a presença e o tamanho dos insertos contidos no vetor p416GPD.

De forma a rastrear mais do que uma colónia por tubo, foram efetuadas reações de PCR com 5 ou 10 colónias da mesma transformação numa única mistura, a fim de detetar a presença do respetivo gene de interesse. As colónias foram também analisadas individualmente no caso de ocorrer a amplificação de fragmentos correspondentes aos genes em estudo. Neste procedimento utilizou-se a enzima Taq II DNA polimerase, bem como os demais reagentes previamente descritos na secção 3.9, referentes à técnica PCR. Foram apenas realizados alguns ajustes específicos nas condições de desnaturação, hibridização e extensão (para detalhes adicionais, consultar os anexos – Tabela A2).

3.10.3 Estratégias alternativas de clonagem

3.10.3.1 Extração de DNAP de um gel de agarose por excisão das bandas:

No primeiro processo testado, o plasmídeo p416GPD foi submetido a uma nova digestão, desta vez utilizando 2 enzimas distintas, a SmaI e a Hind III (ThermoFisher Scientific, USA) (protocolo disponível nos anexos – Tabela A3). Posteriormente, procedeu-se à eletroforese em gel de agarose 1%, visando a extração direta da amostra de plasmídeo digerido do gel, através da excisão das bandas. Neste contexto, foi utilizada a amostra contendo o plasmídeo digerido e uma amostra de plasmídeo não digerido para controlo. Após a migração do gel, a banda resultante do plasmídeo digerido foi cuidadosamente cortada do gel e colocada num tubo de 1,5 mL estéril para se realizar o procedimento de purificação. A purificação do DNAP extraído de um gel de agarose foi realizada utilizando o Kit Gelpure da NZYTech Portugal, de acordo com as instruções do fabricante. A cada 100 mg de gel adicionava-se 300 µL de tampão de ligação, considerando que o peso do gel utilizado tinha 1059,3 mg, foram adicionados 3177 µL de solução tampão. Por fim a amostra de DNAP p416GPD purificada foi guardada a -20 °C para, posteriormente, ser utilizada numa nova clonagem.

Depois da aplicação da técnica mencionada, foi experimentada uma abordagem distinta, em que o plasmídeo p416GPD foi submetido à amplificação por PCR e subsequente digestão utilizando a enzima DpnI.

3.10.3.2 Amplificação do plasmídeo p416GPD por PCR:

O plasmídeo p416GPD foi amplificado na sua totalidade com a enzima *Accuzyme* (Bioline, London UK). A mistura de reação continha 20 ng de p416GPD, os *primers forward* e *reverse* (10 µM) cedidos pelo laboratório (sequências: fwd- GATACCGTCGACCTCG e rev- CTAGTTCTAGAATCCGTC) e *Accuzyme* Mix 2x. O protocolo de amplificação consistiu em apenas 10 ciclos compostos por: desnaturação (95 °C por 30 segundos), hibridização (52 °C por 1 minuto), e extensão (68 °C por 4 minutos e 30 segundos), seguido de uma extensão adicional de 30 segundos a 95 °C. O período de hibridização foi novamente repetido durante um minuto, finalizando com uma fase de extensão a 68°C por 6 minutos e 30 segundos.

O produto da reação final foi incubado com a enzima DpnI, de acordo com as condições especificadas na Tabela 9, para eliminação do DNA molde de p416GPD. Depois de um período de 16h de incubação, o vetor p416GPD foi utilizado para clonagem.

Tabela 9 - Componentes utilizados no processo de digestão de DNA p416GPD com a enzima DpnI.

Componente	Quantidade (µL)
Enzima DpnI (10 U/µL)	1
Tampão	2
Amostra de DNAp	2
H ₂ O U.P.	15
Total	20

3.10.4 Sequenciação dos genes de *C. jadinii* clonados no vetor p416GPD

A confirmação da sequência dos genes de *C. jadinii* inseridos no vetor p416GPD foi efetuada pelo método de Sanger. Para o efeito, 5 µL de cada primer (com uma concentração de 5 pmol/µL) utilizados na amplificação do respetivo gene, foram adicionados a 5 µL da amostra de DNA plasmídico (concentração 100 ng/µL) em tubos diferentes, de modo a obter uma mistura final com um volume total de 10 µL por tubo. Após a preparação das amostras, os tubos foram marcados com o código correspondente e enviados para a *Eurofins Genomics Sequencing*, onde se procedeu à análise de sequenciação.

3.11 Expressão heteróloga dos genes de *C. jadinii* em *S. cerevisiae*

A transformação da levedura *S. cerevisiae* CEN.PK 113-7D com DNA plasmídico foi efetuada de acordo com o protocolo proposto por Gietz e Woods, 2002.

No dia anterior ao ensaio, culturas de células da estirpe *S. cerevisiae* CEN.PK 113-7D foram inoculadas em placas com meio YPD e incubadas a 30 °C. Para a transformação, uma pequena quantidade de biomassa foi ressuspensa em 1 mL de água estéril. A suspensão celular centrifugada durante 1 minuto à velocidade máxima, e o sobrenadante foi descartado. Em paralelo, uma solução de DNA de cadeia simples (SS-DNA, 2 mg/mL), foi incubada a uma temperatura de 100 °C durante 5 minutos num termobloco, e posteriormente em gelo por 2 minutos. Entretanto, ao sedimento contendo as células de *S. cerevisiae*, foram adicionados os seguintes reagentes de forma sequencial: 240 µL de PEG a 50%, 36 µL de Acetato de Lítio (LiAc) 1M, 50 µL de SS-DNA, 6 µL de amostra de DNAp juntamente com 28 µL água ultrapura, num total de 360 µL. A reação foi incubada por um período de 1h a 1h30 num banho térmico a 42 °C. Posteriormente, realizou-se uma centrifugação de 1 minuto, seguida da remoção do sobrenadante. Adicionou-se 500 µL de água estéril e a mistura foi homogeneizada vigorosamente com auxílio de um agitador de tubos. Para concluir, a mistura foi inoculada em placas de YNB glucose e desprovido de uracilo. As placas foram incubadas por 2 a 3 dias a 30 °C, a fim de permitir o crescimento dos transformantes.

3.11.1 Clonagem de genes de *C. jadinii* no vetor p416GDP pelo método de Gap Repair em *S. cerevisiae*

Foi testada uma abordagem distinta em relação ao protocolo supracitado. Nesta abordagem, a clonagem dos genes de *C. jadinii* no plasmídeo p416GPD foi realizada tendo por base o método *GAP Repair*, detalhado na secção 3.10, mas efetuado diretamente na levedura *S. cerevisiae*. Assim, foi elaborada uma mistura de transformação para cada gene, consistindo em 200 *ng* do fragmento de inserção e 100 *ng* de vetor p416GPD. Os fragmentos de inserção, assim como o vetor digerido com a enzima DpnI, foram previamente amplificados por PCR. O restante procedimento de transformação da levedura *S. cerevisiae* é idêntico ao protocolo descrito em 3.11.

3.11.2 Extração de DNA plasmídico de levedura

A extração de DNA plasmídico da levedura *S. cerevisiae* foi efetuada com o Kit NZYMiniprep (NZYTEC, Portugal) utilizando um protocolo adaptado, para confirmação de clones positivos. As quantidades dos tampões A1, A2 e A3 foram duplicadas em relação ao protocolo do fabricante. Após o primeiro tampão ser adicionado na coluna, a solução foi homogeneizada com um agitador de tubos durante 15 minutos na velocidade máxima. Seguidamente, a solução foi incubada por 5 minutos à temperatura ambiente e centrifugada por 2 minutos. O restante protocolo seguiu as instruções fornecidas pelo fabricante. As amostras foram guardadas a -20 °C até posterior utilização.

3.12 Ensaios de crescimento em meio sólido (*drop-tests*)

Foi realizado um ensaio de crescimento em meio sólido com as estirpes de levedura *C. jadinii*, nomeadamente: DSM 2361, PYCC 2598, PYCC 3092, TB104 e TB115 e os transformantes de *S. cerevisiae*. Foram utilizados dois meios de cultura (YP e YNB), suplementados com oito fontes de carbono e energia distintos (ácido acético, cítrico, fumárico, láctico, málico, pirúvico, succínico e glucose - Tabela 4), considerando também duas condições de temperatura (18 °C e 30 °C). O ensaio de crescimento das estirpes de *C. jadinii* nas diferentes fontes de carbono foi realizado em meio

de cultura YP sólido, contrariamente à avaliação do crescimento dos transformantes, o qual foi realizado em meio de cultura YNB sólido.

Com o auxílio de um palito esterilizado, retirou-se uma porção de biomassa dos isolados de culturas frescas e suspendeu-se em 1 mL de água estéril. Após ajuste da suspensão celular para uma DO_{640nm} de 0.2, esta foi submetida a uma série de diluições sucessivas. De cada diluição inoculou-se uma gota de 5 μ L, no sentido da amostra mais diluída para a mais concentrada, a fim de evitar arrastamento de células, no meio desejado. A distribuição das gotas foi efetuada através de uma pipeta multicanal. Após a conclusão do ensaio, as placas foram incubadas em duas condições distintas: a 30 °C durante 4 dias, 18 °C durante 12 dias. Finalizado esse período, procedeu-se ao registo dos resultados através de fotografia.

3.13 Análise bioinformática do genoma de *C. jadinii*

Neste trabalho foram elaboradas análises comparativas *in silico* de 5 estirpes de *C. jadinii*, nomeadamente, PYCC 2578, PYCC 3092, TB104, TB115 e CBS 1600 (estirpe usada como referência) (Tabela 4). Nesse sentido, recolhemos as respetivas sequências genómicas disponíveis no NCBI (*National Center for Biotechnology Information*).

Para todas as estirpes, as sequências (*reads*) foram mapeadas contra o genoma de referência *C. jadinii* CBS 1600, através da ferramenta *Burrows–Wheeler Alignment* (BWA), versão 0.7.17 (-r1188), utilizando os comandos “*index*” e “*mem*” desta ferramenta (Li & Durbin, 2009, 2010). Posteriormente, os dados foram processados utilizando a ferramenta *SAMtools*, versão 1.10, com a execução dos comandos “*view*” e “*sort*” (Tabela 1) (Danecek et al., 2021; Li, 2011; Li et al., 2009).

Para determinar a ploidia das diferentes estirpes em estudo foram executadas funcionalidades do programa estatístico *nQuire*, versão 1.0, utilizando os comandos “*create*” e “*histo*” (Čadež et al., 2021; Weib et al., 2018).

Para a análise de variabilidade genética (*variant calling*), foram utilizados os programas *BCFtools*, versão 1.10.2 (Danecek et al., 2011; Danecek & McCarthy, 2017; Narasimhan et al., 2016), o *Snippy*, versão 4.6.0 e o *Snippy-core*, versão 4.6.0 (Hiltemann et al., 2023; Seemann, 2015).

Numa análise inicial realizada através do *software BCFtools* procedeu-se, separadamente para cada estirpe, à aplicação de cinco comandos de forma sequencial, nomeadamente,

“*mpileup*”, “*call*”, “*norm*”, “*filter*” e “*view*”. Posteriormente, numa fase avançada e após cruzar resultados, foram testadas as funcionalidades dos comandos “*merge*” e “*isec*” (Tabela 2) (Danecek et al., 2011, 2021).

O comando “*filter*” desta ferramenta consiste na filtragem de dados pelo uso de limites fixos, como inclusão, exclusão, restrição de regiões, entre outros, através de parâmetros específicos, como a qualidade do genótipo, “*Qual*” (*genotype quality*) e a cobertura, “*DP*” (*raw read depth*) (<https://samtools.github.io/bcftools/bcftools.html#filter>). Para esta análise procedeu-se à exclusão de dados com baixa qualidade (*Qual* < 20) e baixa cobertura (*DP* < 10) (Silva, 2021). Esta metodologia foi repetida para cada uma das estirpes. Assim, o comando descrito na Figura 1 executa o *software* BCftools, instruindo a seleção dos locais onde as condições selecionadas são aplicadas.

```
~/Desktop$ bcftools filter -Ov -e 'Qual < 20 || DP < 10' -o output.vcf ficheiro_input.norm.vcf
```

Figura 1 - Exemplo modelo da linha de código utilizando a ferramenta BCftools *filter*, com parâmetros definidos de exclusão, qualidade e cobertura do genoma.

Através do comando “*isec*” foi criado um ficheiro que compara os genomas das diferentes estirpes em relação ao genoma de referência, indicando as interseções existentes. Quando a análise abrange 2 ou mais ficheiros a expressão deve ser a seguinte “\$ bcftools isec -p dir -n=4”, em que ‘*n*’ corresponde ao número total de ficheiros a analisar e ‘-p dir’ o nome do diretório onde os resultados serão armazenados (<https://samtools.github.io/bcftools/bcftools.html#isec>). A Figura 2 demonstra a linha de código completa.

```
~/Desktop$ bcftools isec -p dir -n=4 PYCC2578.filter.vcf.gz PYCC3092.filter.vcf.gz  
TB104.filter.vcf.gz TB115.filter.vcf.gz
```

Figura 2 - Exemplo da linha de código utilizando a ferramenta BCftools *isec* para interseção dos ficheiros das quatro estirpes. ‘gz’ indica que os ficheiros estão num formato zipado.

Por último, a abordagem realizada com a ferramenta *Snippy* consistiu na seleção do genoma de uma das estirpes de *C. jadinii* e no genoma de referência para analisar a variabilidade genética existente (<https://github.com/tseemann/snippy>). Foram escolhidos parâmetros avançados como “*minimum mapping quality*” (20) e “*minimum coverage*” (10), tal como aplicado no comando *filter*

mencionado acima. Os *outputs* selecionados, ficheiro zipado e formato de tabela, foram necessários para o passo seguinte. Esta metodologia foi repetida para cada estirpe.

Os ficheiros zipados adquiridos no procedimento anterior foram essenciais para a análise através do *Snippy-core*. Nesta ferramenta foram carregados, simultaneamente, todos os ficheiros no formato zipado pertencentes a cada uma das estirpes em estudo e o genoma de referência no formato FASTA. Os *outputs* escolhidos foram “*tab-separated columnar list of core SNP sites*” e “*tab-separated columnar list of alignment/core-size statistics*” (Hiltemann et al., 2023).

4. ANÁLISE E DISCUSSÃO DOS RESULTADOS

A levedura *Cyberlindnera jadinii* possui várias características que a tornam uma promissora fábrica celular microbiana, e onde se incluem a utilização de ácidos carboxílicos como única fonte de carbono e energia. Este estudo teve o objetivo de aprofundar o conhecimento sobre a fisiologia, metabolismo e genoma desta levedura. Para o efeito foram desenvolvidos vários ensaios, incluído a caracterização fisiológica de diferentes estirpes, a clonagem de transportadores de ácidos carboxílicos putativos e sua expressão heteróloga em *S. cerevisiae*. Dado que foi reportada a existência de uma grande variabilidade na ploidia desta levedura, foi efetuada uma análise comparativa dos genomas das diferentes estirpes de *C. jadinii* sequenciadas.

4.1 Amplificação de genes da levedura *Cyberlindnera jadinii*

Neste trabalho foram selecionados para estudo oito genes de *C. jadinii* (Tabela 10). Estes foram identificados em trabalhos anteriores do grupo de investigação, onde foi efetuada uma análise do transportoma de *C. jadinii*. Estes genes ou possuem homologia com transportadores de ácidos carboxílicos, ou apresentam domínios conservados de famílias de transportadores de ácidos carboxílicos (Sousa-Silva, 2021).

Tabela 10 - Genes de *C. jadinii* selecionados para estudo neste trabalho. Domínios transmembranares (TMSs): PM – membrana plasmática.

Nome / número de acesso	Função anotada	Função prevista	Localização subcelular prevista	Tamanho (bps)	Família	Topologia (TMSs)
CT5 <i>XP_0200677</i> 36.1	<i>MFS general substrate transporter</i>	<i>monocarboxylate transporter; mch</i>	PM	487	MCT	11
CT6 <i>XP_0200686</i> 15.1	<i>hypothetical protein CYBJADRAFT_131197</i>	<i>Dicarboxylic amino acid (L-glutamate and L-aspartate) permease</i>	PM	524	YAT	11
CT7 <i>XP_0200681</i> 61.1	<i>dityrosine transporter A Q resistance</i>	<i>dityrosine transporter</i>	?	488	MFS	12
CT10 <i>XP_0200702</i> 34.1	<i>DUF1212-domain-containing protein</i>	<i>Putative threonine/serine exporter</i>	-	648	-	9

CT11 XP_0200703 72.1	<i>amino acid permease</i>	<i>L-glutamate and L-aspartate transporter</i>	PM	581	SLC	12
CT12 XP_0200686 37.1	<i>hypothetical protein</i> CYBJADRAFT_164167	<i>Sulphate permease</i>	-	520	SulP	6
CT13 XP_0200690 05.1	<i>transmembrane protein</i>	ATO2	PM	280	AceTr	6
CT14 XP_0200688 27.1	<i>MFS general substrate transporter</i>	<i>Quinidine resistance protein</i>	PM	624	DHA1	12

Foram testadas várias reações de PCR para amplificar os genes de *C. jadinii* (Tabela 8). Como descrito na secção 3.9 dos materiais e métodos, inicialmente foi realizada uma mistura de reação utilizando a enzima Taq II DNA polimerase (NZYTech, Portugal), o DNAg de *C. jadinii* DSM 2361, o par de *primers* respetivos a cada gene (Tabela 7) e os restantes reagentes conforme indicado na Tabela 8. Após a extração do DNAg de *C. jadinii* DSM 2361, efetuada através do método de Lööke et al. (2011), obteve-se DNA genómico com uma concentração de 50.2 ng/μl. A temperatura de hibridização para o PCR foi escolhida de acordo com as temperaturas de desnaturação dos *primers* de cada gene, indicadas na Tabela 7, por esse motivo foi criado um gradiente de temperatura para ajustar às condições ideais de cada gene (Jacobus & Gross, 2015). Após os primeiros ensaios de amplificação constatou-se que: os genes CT6, CT11, CT13 e CT14 exibiram baixo rendimento, evidenciando uma banda pouco intensa; o gene CT12 não amplificou corretamente, pois o fragmento observado tem um tamanho superior ao esperado; o gene CT5 não amplificou. As amplificações dos genes CT7 e CT10 ocorreram como o esperado, tendo o tamanho pretendido, sendo que o CT7 demonstrou uma banda mais intensa (Figura 3A).

De acordo com estes resultados foi alterada a temperatura de hibridização no programa de PCR para cada gene (exceção do gene CT7), com um incremento de temperatura para aumentar a especificidade. Esta modificação levou a um aumento da intensidade da banda, particularmente nos genes CT10 e CT11. No entanto, foi ainda necessária a otimização dos parâmetros para os restantes (CT5- sem amplificação, CT6 e CT12- possuíam bandas extra, CT13 e CT14- baixo rendimento) (Figura 3B).

As condições de amplificação foram mais uma vez ajustadas, incluindo um aumento da temperatura de hibridização para 54 °C, aplicado de forma uniforme a todos os genes. Neste

cenário, o gene CT5 amplificou, mas com baixo rendimento, os genes CT13 e CT14 mantiveram baixo rendimento, enquanto o CT6 demonstrou a amplificação desejada. Como controlo positivo foi utilizado o gene CT7 (Figura 3C).

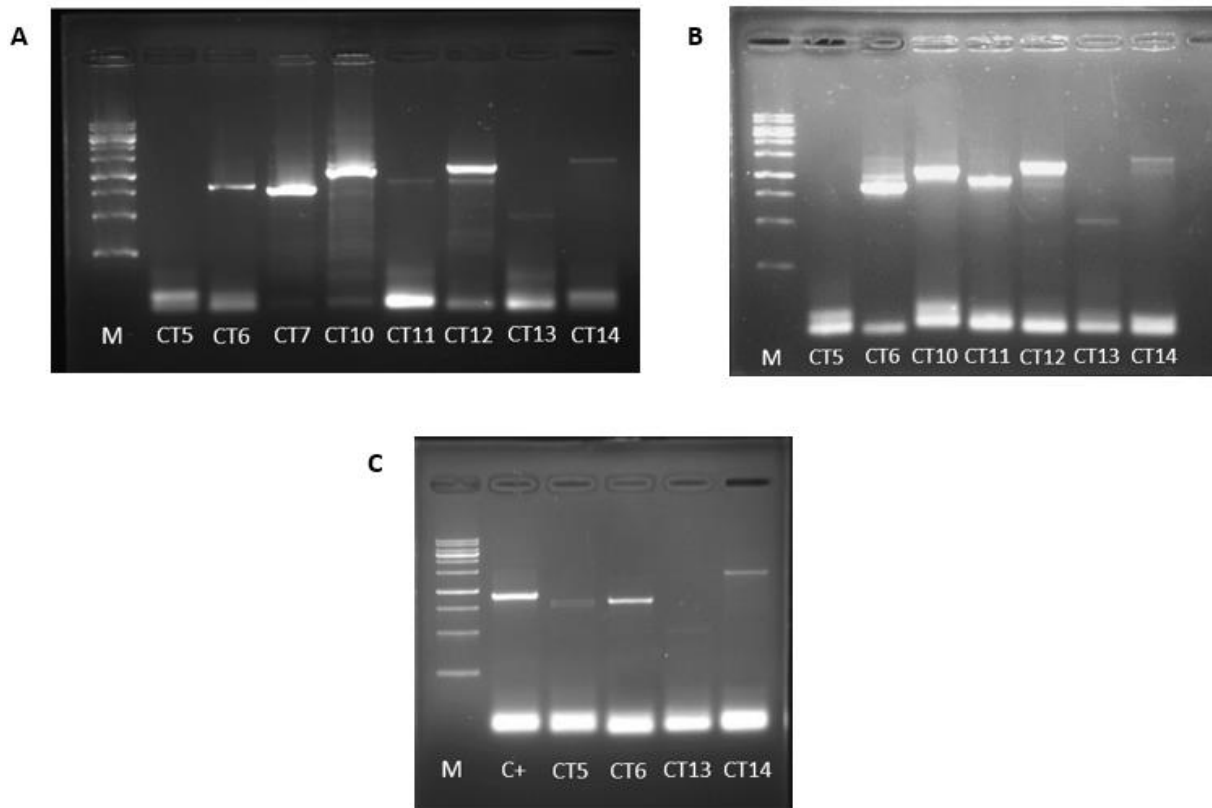


Figura 3 - Produtos de PCR da amplificação de genes de *C. jadinii* com a enzima Taq II. **A)** Amplificação dos genes CT7 e CT10 com gradiente de temperatura na etapa de hibridização. Genes CT5, CT6, CT11, CT12, CT13 e CT14 não amplificaram corretamente. **B)** Amplificação dos genes CT10, CT11 com otimização dos parâmetros. Genes CT5, CT6, CT12, CT13 e CT14 não amplificaram corretamente. **C)** Amplificação dos genes CT5, CT6 e CT14 com alteração da temperatura de hibridização. Gene CT13 não amplificou. M - Marcador molecular: GRS Ladder 1kb (ThermoFisher Scientific); C+ - controlo positivo (gene CT7). Visualizado num gel de agarose 1% sob luz UV.

Após a otimização das condições de amplificação, efetuou-se a amplificação dos genes com a enzima *Accuzyme* DNA polimerase (NZYTech, Portugal), seguindo a reação detalhada na Tabela 8. Esta enzima possui atividade *proofreading*, reduzindo a taxa de erro no decorrer da amplificação de DNA. Conforme demonstrado na Figura 4, observa-se um aumento do rendimento de amplificação para todos os genes ao utilizar a enzima *Accuzyme*, com uma temperatura de hibridização de 52 °C, à exceção do gene CT13. Após vários ensaios e exploração de diferentes parâmetros de amplificação, que incluíram aumento de temperatura de hibridização e alteração das enzimas utilizadas, o gene CT13 continuou a não apresentar amplificação. Dado que a levedura *C. jadinii* apresenta uma elevada variabilidade genética, como será discutido neste trabalho, este

gene poderá ter uma sequência diferente da contida nos *primers*. Assim, em estudos futuros, será necessário considerar a utilização de novos *primers*. A enzima *Accuzyme* tinha sido utilizada anteriormente em ensaios de clonagem de genes que codificam transportadores, particularmente de genes de *C. jadinii* (Soares-Silva et al., 2015; Sousa-Silva, 2021).

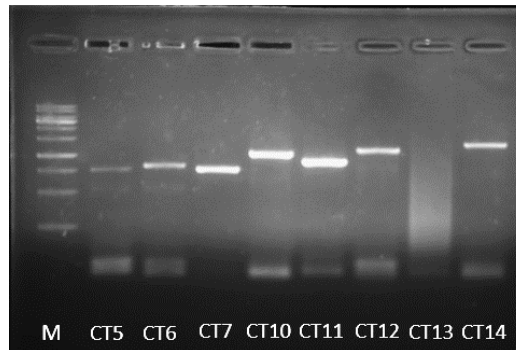


Figura 4 - Produtos de PCR da amplificação de genes de *C. jadinii* com a enzima *Accuzyme*. Amplificação de todos os genes, com exceção do gene CT13. M - Marcador molecular: GRS Ladder 1kb (ThermoFisher Scientific). Visualizado num gel de agarose 1% sob luz UV.

4.2 Clonagem de genes de *C. jadinii* no vetor p416GPD

Para a clonagem dos genes de *C. jadinii* foi selecionado o plasmídeo centromérico p416GPD, que contém um promotor constitutivo (GPD), e que possibilitou a expressão heteróloga de transportadores de carboxilatos em *S. cerevisiae* (Sousa-Silva et al., 2022).

A clonagem de genes de *C. jadinii* foi efetuada pelo método Gap Repair. Após se obter a amplificação dos genes de *C. jadinii*, o objetivo seguinte foi linearizar o plasmídeo p416GPD no local de clonagem múltipla, onde decorrerá a inserção dos genes em estudo e, posteriormente, purificá-lo. Para o efeito utilizou-se a enzima de restrição *EcoRI* (ThermoFisher Scientific, USA) (Jacobus & Gross, 2015). A eficiência da digestão foi determinada por meio de eletroforese em gel de agarose 1% (Figura 5).

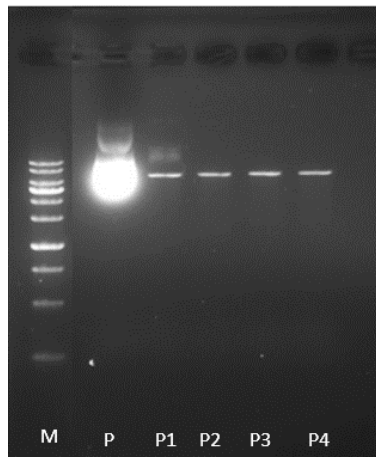


Figura 5 - Digestão do plasmídeo p416GPD com a enzima *EcoRI*. M - Marcador molecular: GRS Ladder 1kb (ThermoFisher Scientific); P - controlo, plasmídeo não digerido; P1 a P4 - quatro amostras diferentes de plasmídeo digerido. Visualizado num gel de agarose 1% sob luz UV.

Para proceder à clonagem dos genes de *C. jadinii* no vetor p416GPD, os produtos de PCR respetivos à amplificação de cada gene foram purificados e, por fim, quantificados. Segundo o método Gap Repair (Jacobus & Gross, 2015), foi elaborada uma mistura com 200 *ng* do fragmento de inserção e 100 *ng* do vetor p416GPD, para cada gene. Esta mistura foi transformada posteriormente em células competentes de *Escherichia coli* XL-1 Blue, as quais preparadas utilizando o protocolo *Rubidium Chloride* (RbCl₂). As células foram plaqueadas em meio de cultura LB suplementado com ampicilina e incubadas a 37 °C, durante a noite.

No dia seguinte realizou-se a extração do DNA plasmídico p416GPD de duas colónias diferentes para cada uma das transformações, ou seja, para cada um dos genes. A determinação da inserção dos genes no vetor p416GPD foi efetuada através da amplificação por PCR. Na Figura 6A é possível verificar que em apenas uma das amostras houve inserção do gene no vetor p416GPD, nomeadamente o gene CT5 (colónia 2).

Para confirmação dos resultados obtidos, as amostras foram sujeitas a digestão com as enzimas de restrição *KpnI* e *XbaI* (ThermoFisher Scientific, USA). Esta digestão foi efetuada para cada clone a fim de verificar se efetivamente houve inserção do gene no vetor. Como se pode observar na Figura 6B, os resultados obtidos são idênticos aos anteriores, constatando-se apenas um clone positivo.

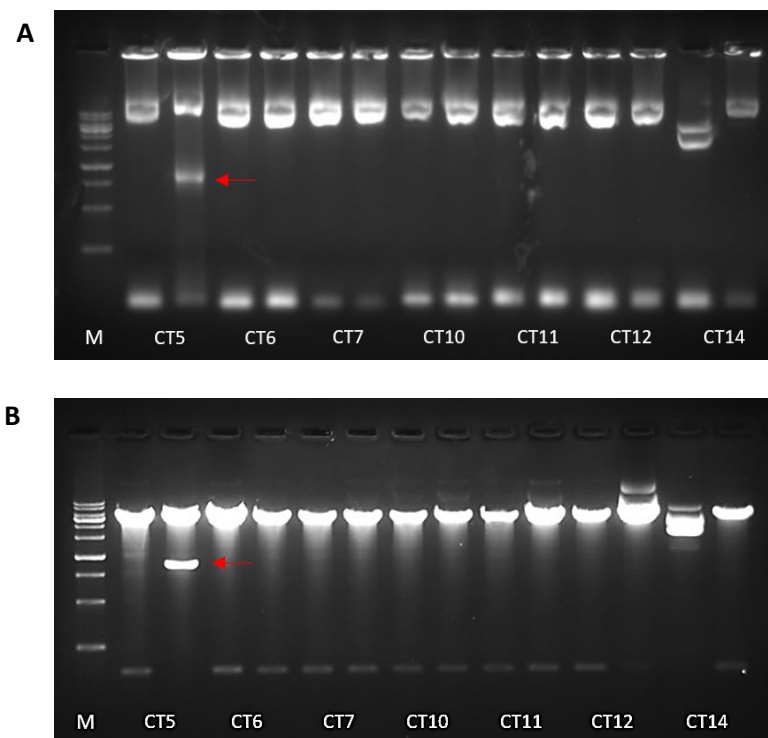


Figura 6 - Clonagem dos genes CT5, CT6, CT7, CT10, CT11, CT12 e CT14 de *C. jadinii* no vetor p416GPD em células de *E. coli* XL-1 Blue. **A)** Amplificação do gene CT5 (colónia 2) clonado no vetor p416GPD. M - Marcador molecular: GRS Ladder 1kb (ThermoFisher Scientific). **B)** Digestão das amostras com as enzimas KpnI e XbaI. M - Marcador molecular: GRS Ladder 1kb (ThermoFisher Scientific). Visualizado num gel de agarose 1% sob luz UV.

De acordo com os resultados anteriores, foram selecionadas outras colónias, para realizar novo um ensaio de PCR em colónia com o objetivo de descobrir novas inserções do gene no vetor p416GPD. Este procedimento envolveu a agregação de várias colónias de cada uma das transformações numa única mistura, e foi replicado para cada gene (à exceção do transformante p416GPD+CT5 onde já se tinha confirmado a inserção). A reação foi efetuada com a enzima Taq II DNA polimerase (NZYTech, Portugal) e o restante protocolo foi realizado como referido na secção 3.10.2 da metodologia. Porém, este método demonstrou o mesmo que os anteriores, concluindo-se que o processo de clonagem não foi eficiente, resultando apenas num clone positivo (p416GPD+CT5 – colónia 2). Este mesmo clone foi usado como controlo positivo nos procedimentos seguintes.

Com o intuito de otimizar os resultados derivados dos procedimentos de clonagem, foram implementadas duas novas abordagens. A primeira consistiu na digestão do plasmídeo p416GPD com as enzimas de restrição SmaI e Hind III (ThermoFisher Scientific, USA) e posterior extração do plasmídeo diretamente do gel de agarose 1%, de forma a evitar a religação do plasmídeo quando transformado em células competentes de *E. coli*. Os genes de *C. jadinii* em estudo foram, mais

uma vez, amplificados por PCR utilizando o DNAg da estirpe DSM 2361 de *C. jadinii*. Foi efetuada uma nova transformação em que se duplicou a quantidade de vetor a utilizar na reação para 200 ng, e os restantes cálculos foram feitos através da plataforma online *nebiocalculator* (<https://nebiocalculator.neb.com/#!/ligation>), para testar diferentes rácios vetor:inserto. Esta mistura foi transformada em células competentes de *Escherichia coli* XL-1 Blue. Contudo, não se obteve colónias para nenhuma das transformações. Presumivelmente o rácio utilizado entre inserto e vetor não foi o adequado. Outra razão para a baixa eficiência da transformação pode estar relacionada com o tamanho do plasmídeo utilizado. O plasmídeo p416GPD contém 5600 pb e alguns dos genes contêm entre 1800 e 2000 pb. Isto faz com que o plasmídeo recombinante tenha um tamanho que possa prejudicar a eficiência da transformação. No entanto, este plasmídeo foi utilizado com sucesso para a clonagem de outros genes da levedura *C. jadinii* (Sousa-Silva et al., 2022).

A segunda abordagem consistiu na amplificação por PCR do plasmídeo p416GPD e posterior digestão com a enzima de restrição DpnI (Jacobus & Gross, 2015), conforme indicado na Tabela 9, para obtenção do plasmídeo linearizado. Procedeu-se então à clonagem e posterior transformação com as mesmas condições mencionadas acima. Obteve-se um elevado número de colónias maioritariamente dos clones com os genes CT6, CT7, CT10 e CT11, pelo que se realizou um novo ensaio de PCR em colónia e análise por eletroforese. Apesar do número elevado de colónias, verificou-se que nenhuma possuía o gene de interesse inserido, tendo apenas amplificado o controlo positivo obtido anteriormente (dados não apresentados).

Seguindo a clonagem bem-sucedida do gene CT5 no vetor p416GPD, este foi sequenciado (*Eurofins Genomics Sequencing*). A sequência analisada não apresentou erros. Deste modo, procedeu-se à expressão heteróloga do gene CT5 em células de levedura *Saccharomyces cerevisiae* CEN.PK 113-7D, para efetuar a sua caracterização funcional.

4.3 Transformação em *Saccharomyces cerevisiae* CEN.PK 113-7D

O gene CT5 clonado no vetor p416GPD foi transformado na estirpe *S. cerevisiae* CEN.PK 113-7D, seguindo o método de Gietz & Woods (2002). As células transformadas foram incubadas a 30 °C por 2 a 3 dias, em placas de meio mínimo contendo glucose e desprovido de uracilo. Para confirmação de clones positivos foi executado o procedimento de PCR em colónia e observação dos resultados através de um gel de agarose 1% (Figura 7). Conforme a figura abaixo, verifica-se que esta transformação foi bem-sucedida com amplificação das duas amostras do clone transformado em *S. cerevisiae*. Assim, o gene CT5 foi caracterizado fisiologicamente, como será descrito na secção 4.5.

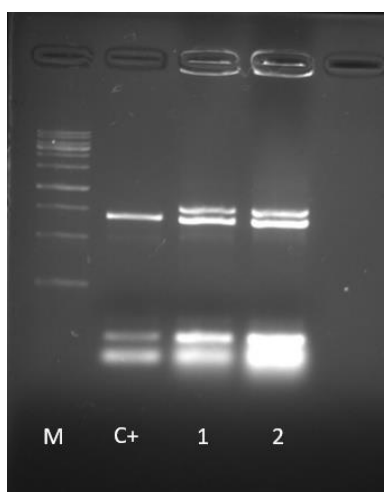


Figura 7 – Confirmação da transformação do gene CT5 clonado no vetor p416GPD na levedura *S. cerevisiae* CEN.PK 113-7D. M - Marcador molecular: GRS Ladder 1kb (ThermoFisher Scientific); C+ - controlo positivo, gene CT5 clonado no vetor p416GPD; 1 – Primeira amostra do clone transformado; 2 – Segunda amostra do clone transformado. Visualizado num gel de agarose 1% sob luz UV.

Para os genes que, anteriormente, foram inseridos no vetor p416GPD a partir da transformação em *E. coli* XL-1 Blue, repetiu-se o procedimento, como mencionado na secção 3.11. Executou-se a inserção dos genes no vetor p416GPD pelo método “*Gap Repair*” na estirpe *S. cerevisiae* CEN.PK 113-7D (Jacobus & Gross, 2015). Desta tentativa resultou uma única colónia pertencente ao gene CT6. A fim de se confirmar e observar a inserção do fragmento e o tamanho do mesmo foi elaborado um ensaio de PCR em colónia e eletroforese em gel de agarose (Figura 8). A Figura 8 demonstra uma possível clonagem bem-sucedida do gene CT6 no vetor p416GPD em *S. cerevisiae* CEN.PK 113-7D, dado que tem o tamanho esperado. Assim, procedeu-se à extração do DNAp utilizando o Kit NZYMiniprep adaptado para leveduras (descrito na secção 3.11.2), para posteriormente sequenciar a amostra.

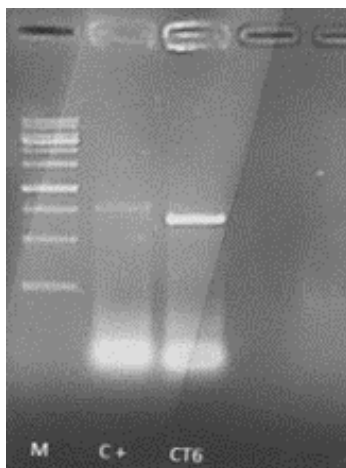


Figura 8 – Confirmação da transformação do gene CT6 clonado no vetor p416GPD em *S. cerevisiae* CEN.PK 113-7D. M - Marcador molecular: GRS Ladder 1kb (ThermoFisher Scientific); C+ - controlo positivo, gene CT5 clonado no vetor p416GPD; CT6 – gene clonado no vetor transformado em levedura. Visualizado num gel de agarose 1% sob luz UV.

Após a obtenção do DNAp extraído, foi realizada a sequenciação da amostra. No entanto, os resultados não foram os esperados, uma vez que o gene não foi sequenciado corretamente. Devido a essa limitação, esta análise ficou interrompida.

As abordagens de transformação testadas neste trabalho não foram 100% eficazes para a maioria dos genes de *C. jadinii*, pelo que novas metodologias deverão ser estudadas mais detalhadamente em trabalhos futuros.

4.4 Caracterização fisiológica de *Cyberlindnera jadinii*

Com o objetivo de perceber o comportamento fenotípico de diversas estirpes de *C. jadinii* em meios suplementados com ácidos carboxílicos como única fonte de carbono e energia e, assim, realizar uma caracterização fisiológica desta levedura, foram realizados ensaios de crescimento em placa. Assim, neste estudo, foi realizado um ensaio de crescimento envolvendo os isolados da levedura *C. jadinii* mencionados na Tabela 3, nomeadamente PYCC 2598, PYCC 3092, TB104, TB105 e como referência, a estirpe DSM 2361. Estes cresceram em meios sólidos YP suplementados com diferentes ácidos carboxílicos, tais como os ácidos acético, láctico, pirúvico, málico, succínico, fumárico, cítrico e glucose, como fonte de carbono e energia. As placas foram incubadas em duas temperaturas diferentes, 18 °C e 30 °C, durante 12 e 4 dias, respetivamente (Figura 9 e 10).

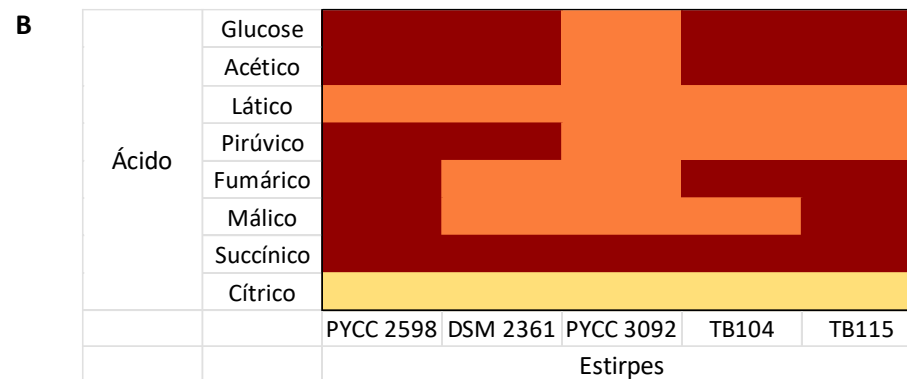
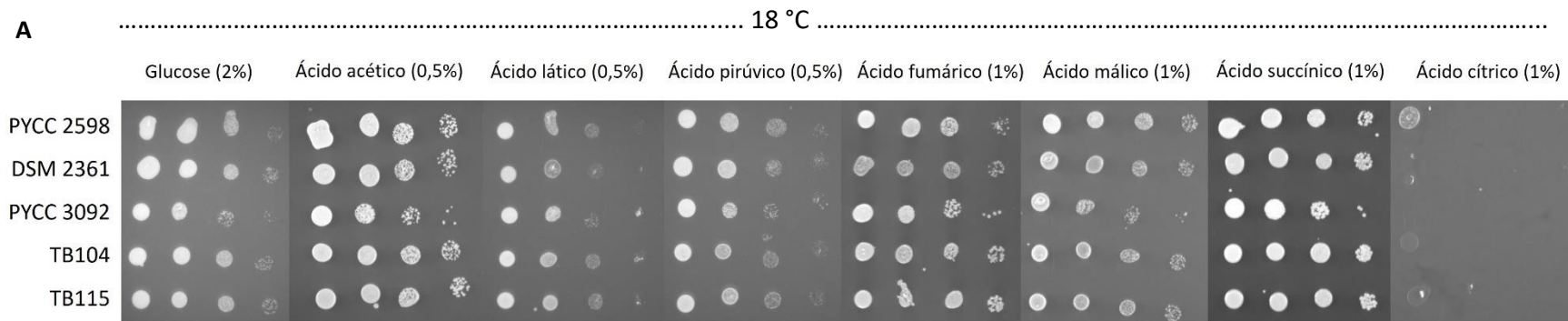


Figura 9 - Padrão de crescimento de cinco estirpes de *C. jadinii* a 18 °C. **A)** Perfis de assimilação das estirpes em meio YP suplementado com glucose (2% p/v), ácido acético (0.5% v/v), ácido láctico (0.5% v/v), ácido pirúvico (0.5% p/v), ácido fumárico (1% p/v), ácido málico (1% p/v), ácido succínico (1% p/v) e ácido cítrico (1% p/v) como fonte de carbono e energia. **B)** Heatmap que representa a capacidade de cada estirpe assimilar a fonte de carbono exibida em A. A cor representa o nível de assimilação: vermelho= crescimento elevado; laranja = crescimento moderado e amarelo= sem crescimento utilizando uma fonte de carbono em particular. As células foram diluídas em série; gotas de 5 µl de cada diluição foram colocadas nas placas, que depois foram incubadas durante 12 dias a 18 °C.

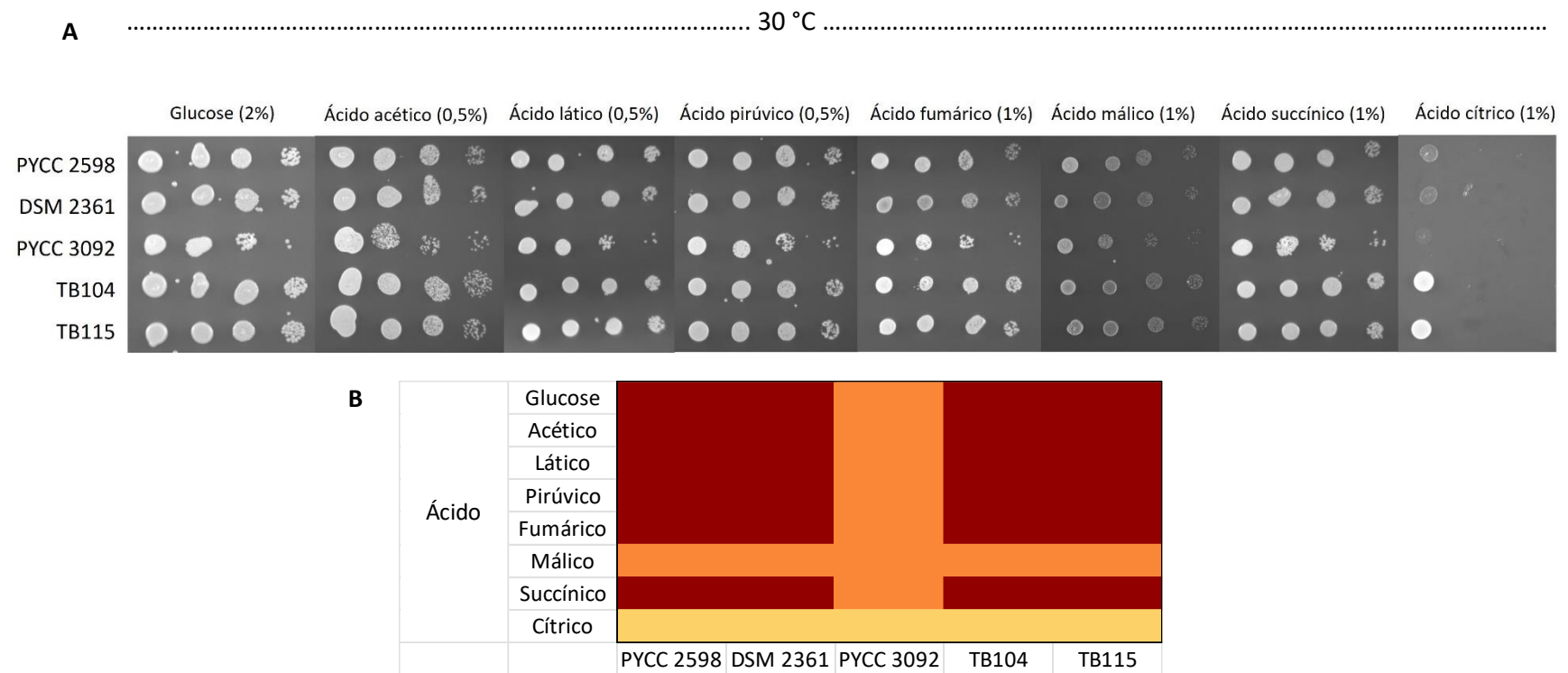


Figura 10 - Padrão de crescimento de cinco estirpes de *C. jadinii* a 30 °C. **A)** Perfis de assimilação das estirpes em meio YP suplementado com glucose (2% p/v), ácido acético (0.5% v/v), ácido láctico (0.5% v/v), ácido pirúvico (0.5% p/v), ácido fumárico (1% p/v), ácido málico (1% p/v), ácido succínico (1% p/v) e ácido cítrico (1% p/v) como fonte de carbono e energia. **B)** Heatmap que representa a capacidade de cada estirpe assimilar a fonte de carbono exibida em A. A cor representa o nível de assimilação: vermelho= crescimento elevado; laranja = crescimento moderado e amarelo= sem crescimento utilizando uma fonte de carbono em particular. As células foram diluídas em série; gotas de 5 µl de cada diluição foram colocadas nas placas, que depois foram incubadas durante 4 dias a 30 °C.

Todas as estirpes de *C. jadinii* apresentaram capacidade de crescimento nas diferentes fontes de carbono, quando cultivadas a 18 °C (Figura 9A). No entanto observa-se uma notável diferença no crescimento no meio suplementado com o ácido cítrico, onde apenas a estirpe PYCC 2598 apresenta um crescimento residual. No meio contendo ácido láctico, as estirpes de *C. jadinii* apresentaram padrões de assimilação menos eficazes em todos os isolados (Figura 9B).

À semelhança do ensaio de crescimento a 18 °C, todas as estirpes foram capazes de assimilar favoravelmente as diferentes fontes de carbono a 30 °C, com exceção do ácido cítrico, onde apenas se verificou crescimento nas estirpes TB104 e TB105, e apenas na gota com maior densidade celular. Neste ensaio de crescimento a 30 °C, as células apresentam um ligeiro decréscimo de crescimento em meios com ácido málico (Figura 10B), porém o ácido láctico foi mais assimilado. De uma forma geral, todas as estirpes apresentaram crescimentos nas fontes de carbono de forma semelhante, tanto a 18°C quanto a 30°C, com crescimento mais rápido na última temperatura (Figura 9 e 10).

Embora o padrão de crescimento das várias estirpes nos diferentes meios seja semelhante, existe uma ligeira diferença no comportamento fenotípico da estirpe PYCC 3092, que apresenta níveis de crescimento menores, nas duas temperaturas de incubação testadas (Figura 9B e 10B). Ainda assim, os resultados indicam que *C. jadinii* poderá utilizar ácidos carboxílicos presentes em resíduos industriais como fonte de carbono e energia para a produção de compostos de valor acrescentado (Sousa-Silva, 2021; Sousa-Silva et al., 2021).

4.5 Expressão heteróloga do gene CT5 de *C. jadinii* em *S. cerevisiae*

A estirpe de *S. cerevisiae* CEN.PK 113-7D transformada com o gene CT5 de *C. jadinii* clonado no vetor p416GPD (p416-CT5) foi utilizada para avaliar a capacidade de as células transformadas crescerem com ácidos carboxílicos como únicas fontes de carbono e energia. Este ensaio de crescimento ocorreu em placa com meio sólido YNB, suplementado com ácidos mono-, di- e tricarboxílicos, a uma temperatura de 30 °C. Como observado na Figura 11, as células apresentaram um crescimento acentuado, como esperado, no meio com glucose, assim como nos meios suplementados com os ácidos acético, láctico e pirúvico. Nos ácidos fumárico, málico e succínico o crescimento foi inferior. Não se observou crescimento em meio contendo ácido cítrico (Figura 11B). Este perfil de crescimento está de acordo com estudos anteriores, pois a levedura *S. cerevisiae* só cresce em meios contendo ácidos monocarboxílicos como fonte de carbono e

energia na gama de valores de pH testados, pois apenas possui proteínas transportadoras de membrana plasmática de ácidos monocarboxílicos, e não de di e tricarboxílicos (Soares-Silva et al., 2020). Quando comparamos o crescimento da estirpe *S. cerevisiae* CEN.PK 113-7D p416-CT5 com o controlo, verificamos que esta apresenta um ligeiro decréscimo no crescimento nos meios que contém ácidos carboxílicos. Este fenótipo é por vezes observado quando há expressão de um exportador de ácidos carboxílicos em *S. cerevisiae*, tal como foi observado na expressão heteróloga do transportador de citrato (*CexA*) de *Aspergillus niger* (Alves et al., 2023). Este facto poderá indicar que o CT5, que é homólogo de transportadores de monocarboxilatos, poderá funcionar como um exportador. No entanto, serão necessários outros estudos para confirmar a sua função, tais como avaliar a produção de ácidos carboxílicos nesta estirpe, por HPLC. Também poderão ser efetuados ensaios de transporte com substratos marcados com ^{14}C para determinar a cinética e especificidade do transportador. Para este efeito, a expressão heteróloga em *S. cerevisiae* do gene CT5 deverá ser efetuada numa estirpe sem transportadores nativos de ácidos carboxílicos, *Jen1* e *Ato1*, assim como dos seus homólogos *Ato2* e *Ato3*.

A levedura *Cyberlindnera jadinii* tem um maior crescimento nestas fontes de carbono quando comparada com outras leveduras, tais como *S. cerevisiae*, que apresenta um crescimento muito pobre em meios contendo ácidos di- e tricarboxílicos (Buerth et al., 2016; Sousa-Silva et al., 2021). Por ser um microrganismo tão promissor, um estudo mais aprofundado dos seus transportadores de membrana (importadores e exportadores), e em particular dos genes que codificam transportadores de carboxilato em *C. jadinii* seria essencial para uma melhor compreensão do seu comportamento fenotípico em distintas condições e meios (Sousa-Silva et al., 2022). A expressão de transportadores membranares específicos demonstrou melhorar o importe de ácidos carboxílicos, como por exemplo o *CjAto5*, o qual, num estudo recente, possibilitou o crescimento celular em todos os ácidos mono-, di- e tricarboxílicos testados (Sousa-Silva, 2021).

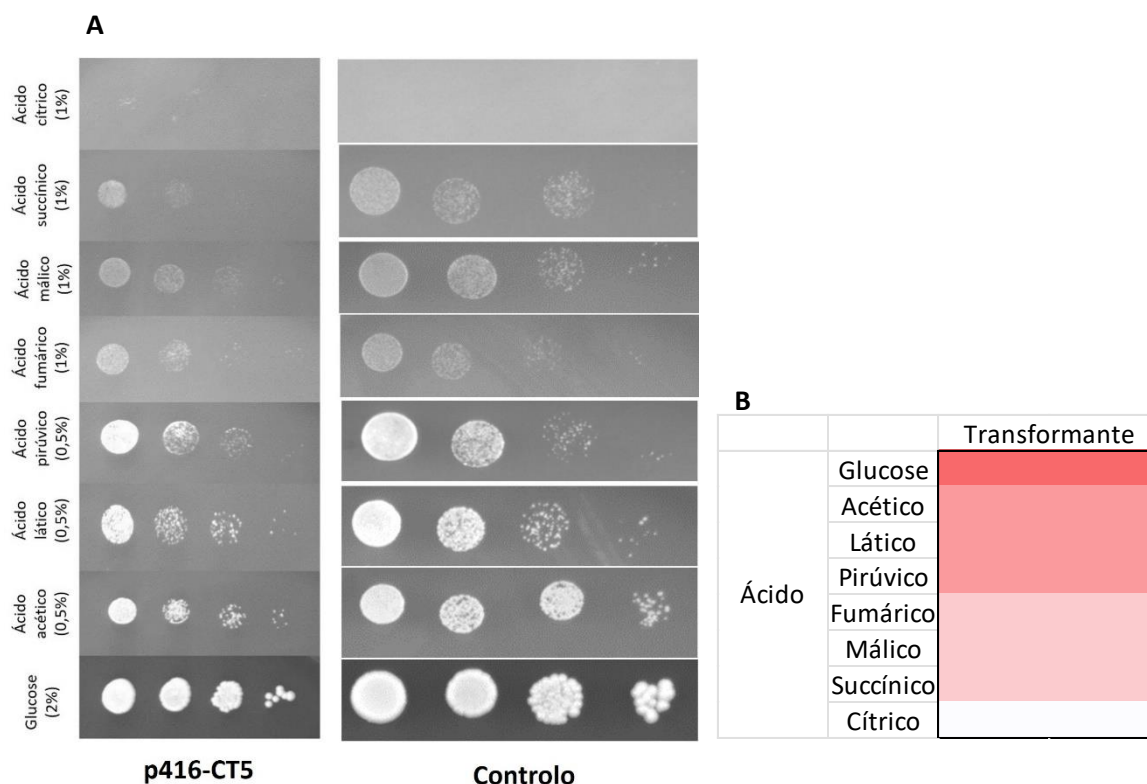


Figura 11 - Fenótipos de crescimento da estirpe *S. cerevisiae* CEN.PK 113-7D transformada com o gene CT5 de *C. jadinii* clonado no vetor p416GPD. **A**) Meios de cultura YNB suplementados com glucose (2% p/v), ácido acético (0.5% v/v), ácido láctico (0.5% v/v), ácido pirúvico (0.5% p/v), ácido fumárico (1% p/v), ácido málico (1% p/v), ácido succínico (1% p/v) e ácido cítrico (1% p/v) testados como única fonte de carbono e energia para o p416-CT5 e o controlo. **B**) Heatmap que representa a capacidade do p416-CT5 assimilar a fonte de carbono exibida em A. A cor representa o nível de assimilação: salmão= crescimento elevado; rosa= crescimento moderado; rosa clarinho= crescimento fraco e branco= sem crescimento utilizando uma fonte de carbono em particular. As células foram diluídas em série; gotas de 5 µl de cada diluição foram colocadas nas placas, que depois cresceram durante 4 dias a 30 °C.

A expressão heteróloga desses transportadores noutras estirpes de *S. cerevisiae* permitirá uma melhor caracterização funcional destes transportadores. Estirpes como a *S. cerevisiae* IMX1000 ou a estirpe *S. cerevisiae* W303-1A *jen1Δ ato1Δ*, demonstraram ser adequadas para a caracterização funcional de transportadores de várias leveduras (Ribas et al., 2017; Soares-Silva et al., 2015; Sousa-Silva, 2021). A estirpe *S. cerevisiae* IMX1000 demonstrou ser eficaz na expressão heteróloga de vários transportadores de *C. jadinii*, dado que apresenta expressão de todos os transportadores carboxílicos conhecidos (Sousa-Silva, 2021; Sousa-Silva et al., 2022). Neste sentido é de extrema importância estudar e caracterizar melhor as proteínas transportadoras codificadas pelas diferentes estirpes desta espécie, em estudos futuros.

Como reportado por diversos autores, *C. jadinii* possui características únicas que a fazem distinguir-se da levedura mais tradicionalmente utilizada, *S. cerevisiae*. As suas características

fisiológicas, metabólicas e genómicas tornam-na uma espécie promissora a ser explorada em aplicações industriais, produção de proteínas recombinantes e metabolitos de interesse comercial, entre outros (Bellut et al., 2019; Cruz, Sterten, et al., 2020; Rupp et al., 2015; Sousa-Silva et al., 2021, 2022). Nas últimas décadas, tem havido um crescente interesse em explorar o potencial comercial e a aplicação de leveduras não convencionais, como a espécie *C. jadinii*, em diversas áreas de pesquisa. Esse interesse tem levado a iniciativas para caracterizar estirpes que possam ser valiosas para aplicações industriais. Essa caracterização envolve a avaliação de características fisiológicas e metabólicas, com o objetivo de identificar estirpes que exibam fenótipos aprimorados e propriedades desejáveis (Sousa-Silva et al., 2022; Sousa-Silva, 2021).

4.6 Análise genómica comparativa

Foi realizada uma análise bioinformática para estudar as características genómicas de *Cyberlindnera jadinii*. Como referido na metodologia, foram utilizados diversos *softwares* para esse efeito, como por exemplo, o *nQuire*, *BWA*, *SAMtools*, *BCFtools*, *Snippy*, entre outros. As abordagens *in silico* consistiram numa avaliação genómica de cinco estirpes de *C. jadinii*, nomeadamente, PYCC 2578, PYCC 3092, TB104, TB115 e CBS 1600 (estirpe usada como referência). Os respetivos genomas já tinham sido previamente sequenciados numa investigação de Sousa-Silva (2021).

4.6.1 Ploidia

Com o objetivo de avaliar a ploidia das diferentes estirpes de *C. jadinii* em estudo os seus genomas foram analisados com o *software nQuire*. De acordo com os resultados obtidos em formato de gráfico podemos constatar que as estirpes PYCC 2578, TB104 e TB115 exibem resultados semelhantes, apresentando genomas diploides. No entanto, a estirpe PYCC 3092 diferencia-se por exibir um genoma triploide (Figura 12). Esta análise foi realizada tendo por base os resultados apresentados numa investigação de Weib et al. (2018). Os resultados ilustrados na Figura 12 estão em conformidade com a informação relatada na literatura, de que vários estados de ploidia foram reportados para diferentes estirpes de *C. jadinii* variando entre haploidia, diploidia, triploidia e tetraploidia (Buerth et al., 2016; Ikushima, Fujii, & Kobayashi, 2009; Rupp et al., 2015). Isto pode ocorrer pela duplicação e/ou deleção de fragmentos. As estirpes de maior ploidia de *C. jadinii* podem ser vantajosas, por exemplo, para inserção de unidades de expressão.

Isso permite a otimização e ajuste dos níveis de produção de proteínas heterólogas (Rupp et al., 2015). Ainda assim estes resultados evidenciam o caráter predominantemente diploide de *C. jadinii*, já documentado na literatura (Buerth et al., 2016; Krahulec et al., 2020; Rupp et al., 2015; Sousa-Silva et al., 2021).

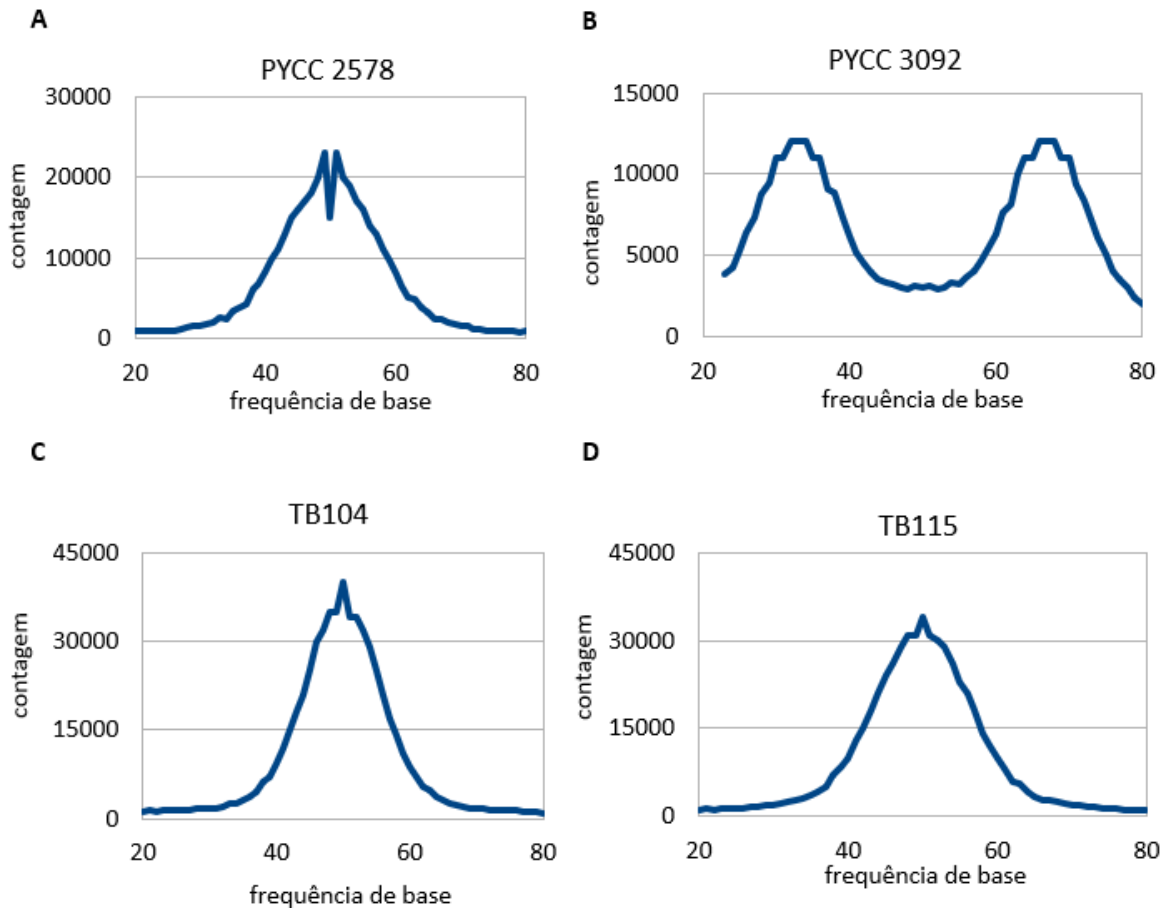


Figura 12 – Análise da ploidia de diferentes estirpes de *C. jadinii* utilizando o programa *nQuire*. Distribuição das frequências de bases em locais variáveis. **A)** estirpe PYCC 2578 diploide. **B)** estirpe PYCC 3092 triploide. **C)** estirpe TB104 diploide. **D)** estirpe TB115 diploide.

4.6.2 Anotação do genoma

De forma a efetuar uma avaliação genómica comparativa entre estirpes foi realizada a anotação dos genomas das estirpes em estudo. Esta análise permitiu detalhar diferenças entre as estirpes, assim como analisar *SNP's* e *indels*, através de ferramentas bioinformáticas. O genoma da estirpe *C. jadinii* CBS 1600 foi utilizado como referência nos procedimentos executados.

Após as sequências genômicas das estirpes serem mapeadas contra o genoma de referência através dos programas *BWA* e *SAMtools* realizou-se a análise de variabilidade genética, utilizando o programa *BCFtools*. Através das funcionalidades do comando “*filter*”, excluiu-se dados de baixa qualidade e baixa cobertura. Assim, obteve-se para cada estirpe de forma individual os locais onde existe variação em relação à referência, assim como a presença de SNP’s e inserções (dados não apresentados).

Visto que era essencial uma análise comparativa que apresentasse os resultados de todas as estirpes simultaneamente, procedeu-se à análise de variabilidade genética através das ferramentas *Snippy* e *Snippy-core*, em conjunto. Estas ferramentas demonstraram a variação relativamente às bases nitrogenadas presentes entre o genoma de cada estirpe contra o genoma de referência (*C. jadinii* CBS 1600). Forneceram ainda, dados sobre a posição da sequência genômica onde essa variação ocorre, deste modo, agruparam-se os resultados de forma a visualizar mais facilmente as maiores diferenças e semelhanças existentes. Através desta abordagem foi possível perceber qual era a estirpe que detinha maiores diferenças comparativamente às outras estirpes. Verificou-se que a estirpe com maior diferença nucleotídica em relação ao genoma de referência era a PYCC 2578, com ~70% da sequência analisada diferente. Consequentemente, esta estirpe também apresentou a maior distinção entre as restantes estirpes analisadas. Este valor carece de confirmação adicional por outros métodos, devido a ser demasiado alto. No entanto, nesta análise estamos a comparar a totalidade dos nucleótidos, em regiões codificantes e não codificantes, sendo de prever que a comparação posterior das sequências aminoacídicas revelará diferenças de uma ordem de grandeza bastante inferior. Por outro lado, a estirpe que mais se identifica neste parâmetro com a referência é a TB104, apresentando apenas ~15% do genoma diferente do genoma de *C. jadinii* CBS 1600 (Figura 13).

Análises complementares realizadas em estudos recentes (Sousa-Silva, 2021) demonstraram que a sequenciação da estirpe PYCC 2578 apresentou um menor número de *reads* e *coverage* em relação a outras estirpes, assim como um tamanho de genoma menor (~18Mbp) em comparação com outras PYCC e TB. Detém ainda um número consideravelmente menor de genes envolvidos na produção e conversão de energia celular, em comparação com as outras estirpes de *C. jadinii* e *S. cerevisiae*. Contrariamente, a PYCC 3092 apresentou o maior número de *reads* e *coverage*, na mesma análise (Sousa-Silva, 2021). Existe ainda pouca informação disponível na literatura sobre estas estirpes de *C. jadinii* em específico, pelo que seria relevante uma análise mais detalhada em estudos futuros.

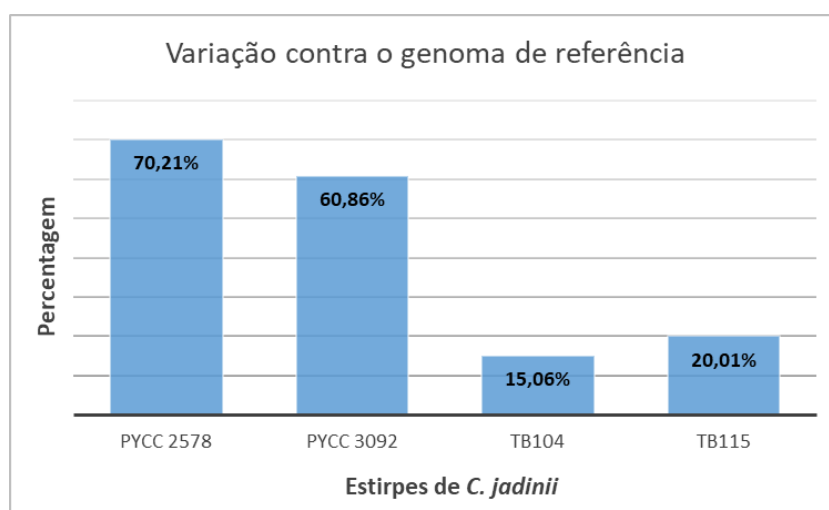


Figura 13 - Variação nucleotídica de estirpes de *C. jadinii* contra o genoma de referência (*C. jadinii* CBS 1600). Análise efetuada com dados provenientes da utilização da ferramenta *Snippy-core*. A estirpe PYCC 2578 apresenta 70,21% de diferença nucleotídica contra a referência, considerando a totalidade do genoma (regiões codificantes e não codificantes); a PYCC 3092 apresenta 60,86% de diferença; a TB104 apresenta apenas 15,06% de diferença; a TB115 apresenta 20,01% de diferença.

Como é possível observar na Tabela 11, com base na análise com a ferramenta *Snippy-core*, verificam-se alguns dos locais de variação do genoma entre as diferentes estirpes e o genoma de referência. Em conformidade com os resultados apresentados na Figura 13, apenas a estirpe PYCC 2578 difere em relação ao genoma de referência, em todas as posições destacadas.

Tabela 11 - Variabilidade genética de estirpes de *C. jadinii* contra a referência *C. jadinii* CBS 1600. Análise comparativa realizada com alguns dos dados resultantes da utilização da ferramenta *Snippy* e *Snippy-core*.

CHROM ^a	Posição	Referência	PYCC 2578	PYCC 3092	TB104	TB115
CDQK01000001.1	26883	C	T	C	C	C
CDQK01000001.1	27102	G	C	G	G	G
CDQK01000001.1	28082	G	A	G	G	G
CDQK01000001.1	28511	G	A	G	G	G
CDQK01000001.1	29061	C	T	C	C	C
CDQK01000001.1	29331	A	G	A	A	A
CDQK01000001.1	29340	G	A	G	G	G
CDQK01000001.1	29343	T	C	T	T	T
CDQK01000001.1	29352	C	T	C	C	C
CDQK01000001.1	29382	A	G	A	A	A

^a - CHROM: significa a sequência em que a variante foi encontrada.

Para uma análise mais completa dos genomas das estirpes empregou-se o comando “*isec*” da ferramenta *BCFtools* com o objetivo de assinalar as correspondências existentes entre estirpes, ao nível dos nucleótidos. Esta análise apenas apresenta as posições em que todas as estirpes em estudo possuem a mesma base. Assim, esta metodologia foi útil para: perceber as interseções entre as estirpes em estudo (as PYCC e TB); em que posição as mesmas ocorrem; observar inserções. Através da Tabela 12 é possível verificar que, por exemplo: na posição 3331 o genoma de referência (*C. jadinii* CBS 1600) possui a base Timina, enquanto todas as outras estirpes em estudo possuem Citosina; nas posições 3358, 3470 e 3661 foram observadas inserções de bases nos genomas das estirpes em relação à referência. Existindo variação/diferença entre qualquer uma das estirpes em estudo numa determinada posição, essa informação já não estará discriminada nesta análise, utilizando o comando “*isec*”.

Tabela 12 - Correspondência genética de estirpes de *C. jadinii* contra o genoma de referência CBS 1600. Análise realizada com alguns dos dados resultantes da utilização da ferramenta *BCFtools*, comando *isec*.

CHROM	Posição	Referência	Correspondência
CDQK01000001.1	3331	T	C
CDQK01000001.1	3344	G	T
CDQK01000001.1	3358	C	CAAG
CDQK01000001.1	3376	C	T
CDQK01000001.1	3383	G	A
CDQK01000001.1	3414	T	C
CDQK01000001.1	3441	G	A
CDQK01000001.1	3457	G	C
CDQK01000001.1	3470	A	AG
CDQK01000001.1	3481	G	T
CDQK01000001.1	3589	C	T
CDQK01000001.1	3633	C	T
CDQK01000001.1	3661	G	GCTT

^a - CHROM: significa a sequência em que a variante foi encontrada.

Deste modo, e de uma forma abrangente, a consulta e análise destes resultados em conjunto dá-nos informação sobre: a estirpe com maior distinção nucleotídica do genoma de referência; a variabilidade genética presente entre as estirpes de *C. jadinii*; as posições do genoma em que todas as estirpes em estudo possuem a mesma base, porém distinta do genoma de referência; a presença de indels (sobretudo inserções) e SNP's; a semelhança entre as estirpes PYCC; semelhança entre as estirpes TB.

A detecção de indels e SNP's nas estirpes de *C. jadinii* reforça a existência de variação genética intraespecífica, que pode ser explorada no sentido de compreender melhor a diversidade genética dentro desta espécie. Estas variações podem ser relevantes na perspectiva de aplicações biotecnológicas, como por exemplo, para a produção de metabolitos de interesse comercial e produção de proteínas heterólogas (Rupp et al., 2015). Podem igualmente ser utilizadas para melhorar estirpes para produção industrial. No entanto, as variações específicas que ocorrem em determinadas posições no genoma podem levar a implicações na adaptação e estabilidade das estirpes em diferentes condições e ambientes. Adicionalmente, variações genómicas identificadas podem influenciar o fenótipo das estirpes, o que pode incluir diferenças na capacidade de crescimento em determinadas fontes de carbono. Curiosamente, a estirpe PYCC 2578 que, neste trabalho, apresentou maior distinção nucleotídica em relação ao genoma de referência é a que possui melhores resultados no crescimento em meios suplementados com diferentes ácidos carboxílicos, de acordo com estudos anteriores (Sousa-Silva, 2021).

Dado o atual déficit de informação na literatura sobre diferentes estirpes de *C. jadinii*, futuramente devem ser desenvolvidas metodologias mais avançadas para análise bioinformática, de forma a obter um conhecimento mais profundo, particularmente, sobre as estirpes de *C. jadinii* em questão. Além disso, a construção e análise de estirpes mutantes podem desempenhar um papel crucial na investigação das características únicas de *C. jadinii*. Os resultados obtidos através destas análises podem auxiliar em futuras investigações.

5. CONCLUSÃO E PERSPETIVAS FUTURAS

A levedura *Cyberlindnera jadinii* desempenha um papel cada vez mais importante na indústria biotecnológica. A sua elevada versatilidade, assim com a capacidade de utilização de uma ampla gama de ácidos orgânicos como única fonte de carbono e energia torna-a uma escolha atrativa para a produção de diversos compostos de interesse comercial. Além disso, a sua robustez e capacidade de crescer em meios de cultura económicos com alta produtividade reforçam a sua importância como uma promissora fábrica celular microbiana para a biotecnologia. Assim, neste trabalho foram estudadas características fisiológicas e genómicas de *C. jadinii* com o objetivo de aprofundar o conhecimento sobre esta levedura e otimizar as suas propriedades como fábrica celular microbiana. O principal foco desta investigação envolveu a caracterização fisiológica de diferentes estirpes de *C. jadinii*, nomeadamente a sua capacidade de utilização de diferentes ácidos carboxílicos, assim como a identificação de novos transportadores de ácidos carboxílicos. Simultaneamente, foi efetuada uma análise genómica comparativa para determinar a variabilidade intra-espécie.

Os estudos efetuados demonstraram que todas as estirpes testadas neste trabalho foram capazes de assimilar ácidos mono e dicarboxílicos. Foi possível verificar que o ácido cítrico foi a fonte de carbono menos utilizada pela generalidade das estirpes, no entanto os isolados selvagens TB104 e TB115 apresentaram um crescimento significativo neste meio. Além disso, constatou-se uma ligeira redução na capacidade de crescimento da estirpe PYCC 3092 em todos os meios testados. Essa análise proporcionou uma melhor compreensão sobre o comportamento de *C. jadinii* em relação à assimilação de ácidos carboxílicos e a sua adaptabilidade a diferentes fontes de carbono e temperaturas. Uma comparação do transportoma destas diferentes estirpes de *C. jadinii* poderá revelar quais os transportadores de ácidos carboxílicos que serão responsáveis pelos diferentes fenótipos observados. Neste trabalho foi caracterizado por expressão heteróloga em *Saccharomyces cerevisiae* o gene CT5, que poderá funcionar como um exportador de ácidos carboxílicos. No entanto serão necessários mais estudos para determinar a função do CT5, incluindo a expressão deste gene numa estirpe de *S. cerevisiae* sem transportadores de ácidos carboxílicos, de forma a avaliar os parâmetros cinéticos do transporte de ácidos carboxílicos usando substratos marcados com ^{14}C . Dado que este gene poderá codificar um exportador, será também importante avaliar a produção de ácidos carboxílicos, usando para o efeito a

determinação da presença destes ácidos no meio de crescimento, por HPLC. Será também importante efetuar a caracterização funcional dos restantes transportadores de ácidos carboxílicos putativos incluídos neste estudo.

Algumas limitações foram identificadas nos processos de clonagem e transformação em *S. cerevisiae*, etapas essenciais para a caracterização fisiológica dos genes de *C. jadinii* em estudo. Deste modo, é imperativo a otimização dessas técnicas. Isso pode incluir a exploração de novos *primers*, avaliação de diferentes estirpes para extração de DNAg ou alteração das quantidades de DNA genómico e plasmídico usadas na estratégia de clonagem.

Relativamente à análise genómica foram detetadas algumas diferenças entre as estirpes em estudo através das ferramentas bioinformáticas utilizadas. A estirpe PYCC 2578 exibiu a maior discrepância em relação à estirpe de referência CBS 1600 e as restantes. Foram também observadas algumas diferenças ao nível da ploidia, particularmente na estirpe PYCC 3092. Trabalhos futuros deverão focar-se em associar estas diferenças com as diferenças observadas a nível fenotípico, estabelecendo possíveis correlações entre a presença/ausência de determinados SNPs, e o conjunto de características fenotípicas das estirpes. Visto que a análise genómica de *C. jadinii* ainda é uma área em desenvolvimento, a disponibilidade de genomas de *C. jadinii* ainda é limitada na literatura científica. Este cenário destaca a necessidade de investigações mais aprofundadas e detalhadas, uma vez que as particularidades dessas estirpes podem conter *insights* valiosos para diversas aplicações. Portanto, estudos futuros envolvendo abordagens *cross-omics*, para análise fenómica, podem fornecer informações cruciais particularmente para a área da biotecnologia.

Em suma, o trabalho apresentado nesta tese evidenciou o potencial biotecnológico da levedura *C. jadinii* e enriqueceu os conhecimentos das suas características fisiológicas e genómicas. Os resultados obtidos poderão auxiliar e ser aplicados em investigações futuras.

6. BIBLIOGRAFIA

- Abuín, J. M., Pichel, J. C., Pena, T. F., & Amigo, J. (2015). BigBWA: Approaching the Burrows-Wheeler aligner to Big Data technologies. *Bioinformatics*, 31(24), 4003–4005. <https://doi.org/10.1093/bioinformatics/btv506>
- Ali, A. H., & Hipkin, C. R. (1986). Nitrate assimilation in *Candida nitratophila* and other yeasts. *Archives of Microbiology*, 144(3), 263–267. <https://doi.org/10.1007/BF00410960>
- Alves, J., Sousa-Silva, M., Soares, P., Sauer, M., Casal, M., & Soares-Silva, I. (2023). Structural characterization of the *Aspergillus niger* citrate transporter CexA uncovers the role of key residues S75, R192 and Q196. *Computational and Structural Biotechnology Journal*, 21, 2884–2898. <https://doi.org/10.1016/j.csbj.2023.04.025>
- Bekatorou, A., Psarianos, C., & Koutinas, A. A. (2006). Production of food grade yeasts. *Food Technology and Biotechnology*, 44(3), 407–415.
- Bellut, K., Michel, M., Zarnkow, M., Hutzler, M., Jacob, F., Atzler, J. J., Hoehnel, A., Lynch, K. M., & Arendt, E. K. (2019). Screening and application of *Cyberlindnera* yeasts to produce a fruity, non-alcoholic beer. *Fermentation*, 5(4). <https://doi.org/10.3390/fermentation5040103>
- Benítez, J. A., Alonso, A., Delgado, J., & Kotyk, A. (1983). Sulphate transport in *Candida utilis*. *Folia Microbiologica*, 28(1), 6–11. <https://doi.org/10.1007/BF02877377>
- Boyarskiy, S., & Tullman-Ercek, D. (2015). Getting pumped: Membrane efflux transporters for enhanced biomolecule production. *Current Opinion in Chemical Biology*, 28, 15–19. <https://doi.org/10.1016/j.cbpa.2015.05.019>
- Boze, H., Moulin, G., & Galzy, P. (1992). Production of food and fodder yeasts. *Critical Reviews in Biotechnology*, 12(1–2), 65–86. <https://doi.org/10.3109/07388559209069188>
- Buerth, C., Heilmann, C. J., Klis, F. M., de Koster, C. G., Ernst, J. F., & Tielker, D. (2011). Growth-dependent secretome of *Candida utilis*. *Microbiology*, 157(9), 2493–2503. <https://doi.org/10.1099/mic.0.049320-0>
- Buerth, C., Tielker, D., & Ernst, J. F. (2016). *Candida utilis* and *Cyberlindnera (Pichia) jadinii*: yeast relatives with expanding applications. *Applied Microbiology and Biotechnology*, 100(16), 6981–6990. <https://doi.org/10.1007/s00253-016-7700-8>
- Čadež, N., Dlačny, D., Tome, M., & Péter, G. (2021). *Novakomyces olei* sp. Nov., the first member of a

novel taphrinomycotina lineage. *Microorganisms*, 9(2), 1–19.
<https://doi.org/10.3390/microorganisms9020301>

Campos, J. M., Stamford, T. L. M., Rufino, R. D., Luna, J. M., Stamford, T. C. M., & Sarubbo, L. A. (2015). Formulation of mayonnaise with the addition of a bioemulsifier isolated from *Candida utilis*. *Toxicology Reports*, 2, 1164–1170. <https://doi.org/10.1016/j.toxrep.2015.08.009>

Casal, M., Paiva, S., Queirós, O., & Soares-Silva, I. (2008). Transport of carboxylic acids in yeasts. *FEMS Microbiology Reviews*, 32(6), 974–994. <https://doi.org/10.1111/j.1574-6976.2008.00128.x>

Cássio, F., & LeñO, C. (1993). A comparative study on the transport of L(-)malic acid and other short-chain carboxylic acids in the yeast *Candida utilis*: Evidence for a general organic acid permease. *Yeast*, 9(7), 743–752. <https://doi.org/10.1002/yea.320090708>

Cruz, A., Håkenåsen, I. M., Skugor, A., Mydland, L. T., Åkesson, C. P., Hellestveit, S. S., Sørby, R., Press, C. M. L., & Øverland, M. (2019). *Candida utilis* yeast as a protein source for weaned piglets: Effects on growth performance and digestive function. *Livestock Science*, 226(March), 31–39. <https://doi.org/10.1016/j.livsci.2019.06.003>

Cruz, A., Sterten, H., Steinhoff, F. S., Mydland, L. T., & Øverland, M. (2020). *Cyberlindnera jadinii* yeast as a protein source for broiler chickens: effects on growth performance and digestive function from hatching to 30 days of age. *Poultry Science*, 99(6), 3168–3178. <https://doi.org/10.1016/j.psj.2020.01.023>

Cruz, A., Tauson, A. H., Matthiesen, C. F., Mydland, L. T., & Øverland, M. (2020). *Cyberlindnera jadinii* yeast as a protein source for growing pigs: Effects on protein and energy metabolism. *Livestock Science*, 231(October 2019), 103855. <https://doi.org/10.1016/j.livsci.2019.103855>

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), 1–4. <https://doi.org/10.1093/gigascience/giab008>

Danecek, P., & McCarthy, S. A. (2017). BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics*, 33(13), 2037–2039. <https://doi.org/10.1093/bioinformatics/btx100>

de Carvalho, L. M., Borelli, G., Camargo, A. P., de Assis, M. A., de Ferraz, S. M. F., Fiamenghi, M. B., José, J., Mofatto, L. S., Nagamatsu, S. T., Persinoti, G. F., Silva, N. V., Vasconcelos, A. A., Pereira, G. A.

- G., & Carazzolle, M. F. (2019). Bioinformatics applied to biotechnology: A review towards bioenergy research. *Biomass and Bioenergy*, 123(March), 195–224. <https://doi.org/10.1016/j.biombioe.2019.02.016>
- Dirihan, S., Terho, P., Helander, M., & Saikkonen, K. (2013). Efficient analysis of ploidy levels in plant evolutionary ecology. *Caryologia: International Journal of Cytology*, 66(3), 251–256. <https://doi.org/10.1080/00087114.2013.849414>
- Entian, K. D., & Kötter, P. (2007). 25 Yeast Genetic Strain and Plasmid Collections. *Methods in Microbiology*, 36(06), 629–666. [https://doi.org/10.1016/S0580-9517\(06\)36025-4](https://doi.org/10.1016/S0580-9517(06)36025-4)
- Etherington, G. J., Ramirez-Gonzalez, R. H., & MacLean, D. (2015). Bio-samtools 2: A package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. *Bioinformatics*, 31(15), 2565–2567. <https://doi.org/10.1093/bioinformatics/btv178>
- Failla, M. L., Benedict, C. D., & Weinberg, E. D. (1976). Accumulation and storage of Zn²⁺ by *Candida utilis*. *Journal of General Microbiology*, 94(1), 23–36. <https://doi.org/10.1099/00221287-94-1-23>
- Ferreira, B. H., Ramírez-Prado, J. H., Neves, G. W. P., Torrado, E., Sampaio, P., Felipe, M. S. S., Vasconcelos, A. T., Goldman, G. H., Carvalho, A., Cunha, C., Lopes-Bezerra, L. M., & Rodrigues, F. (2019). Ploidy determination in the pathogenic fungus *Sporothrix* spp. *Frontiers in Microbiology*, 10(FEB). <https://doi.org/10.3389/fmicb.2019.00284>
- Fonseca, G. G., Heinzle, E., Wittmann, C., & Gombert, A. K. (2008). The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Applied Microbiology and Biotechnology*, 79(3), 339–354. <https://doi.org/10.1007/s00253-008-1458-6>
- Gavrilescu, M., & Chisti, Y. (2005). Biotechnology - A sustainable alternative for chemical industry. *Biotechnology Advances*, 23(7–8), 471–499. <https://doi.org/10.1016/j.biotechadv.2005.03.004>
- Gietz, R. D., & Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in Enzymology*, 350(2001), 87–96. [https://doi.org/10.1016/S0076-6879\(02\)50957-5](https://doi.org/10.1016/S0076-6879(02)50957-5)
- Gong, Z., Nielsen, J., & Zhou, Y. J. (2017). Engineering Robustness of Microbial Cell Factories. *Biotechnology Journal*, 12(10). <https://doi.org/10.1002/biot.201700014>
- Gu, L., Zhang, R., Fan, X., Wang, Y., Ma, K., Jiang, J., Li, G., Wang, H., Fan, F., & Zhang, X. (2023). Development of CRISPR/Cas9-Based Genome Editing Tools for Polyploid Yeast *Cyberlindnera jadinii* and Its Application in Engineering Heterologous Steroid-Producing Strains. *ACS Synthetic Biology*, 12(10), 2947–2960. <https://doi.org/10.1021/acssynbio.3c00278>

- Heux, S., Meynial-Salles, I., O'Donohue, M. J., & Dumon, C. (2015). White biotechnology: State of the art strategies for the development of biocatalysts for biorefining. *Biotechnology Advances*, 33(8), 1653–1670. <https://doi.org/10.1016/j.biotechadv.2015.08.004>
- Hiltemann, S., Rasche, H., Gladman, S., Hotz, H. R., Larivière, D., Blankenberg, D., Jagtap, P. D., Wollmann, T., Bretaudeau, A., Goué, N., Griffin, T. J., Royaux, C., Bras, Y. Le, Mehta, S., Syme, A., Coppens, F., Droesbeke, B., Soranzo, N., Bacon, W., ... Batut, B. (2023). Galaxy Training: A powerful framework for teaching! *PLoS Computational Biology*, 19(1), 1–18. <https://doi.org/10.1371/journal.pcbi.1010752>
- Ikushima, S., Fujii, T., & Kobayashi, O. (2009). Efficient gene disruption in the high-ploidy yeast *Candida utilis* using the Cre-loxP system. *Bioscience, Biotechnology and Biochemistry*, 73(4), 879–884. <https://doi.org/10.1271/bbb.80799>
- Ikushima, S., Fujii, T., Kobayashi, O., Yoshida, S., & Yoshida, A. (2009). Genetic engineering of *Candida utilis* yeast for efficient production of L-lactic acid. *Bioscience, Biotechnology and Biochemistry*, 73(8), 1818–1824. <https://doi.org/10.1271/bbb.90186>
- Jacobus, A. P., & Gross, J. (2015). Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli*. *PLoS ONE*, 10(3), 1–17. <https://doi.org/10.1371/journal.pone.0119221>
- Kaikabo, A. A., & Kalshingi, H. A. (2008). Concepts Of Bioinformatics And Its Application In Veterinary Research And Vaccines Development. In *Nigerian Veterinary Journal* (Vol. 28, Issue 2). <https://doi.org/10.4314/nvj.v28i2.3554>
- Kell, D. B., Swainston, N., Pir, P., & Oliver, S. G. (2015). Membrane transporter engineering in industrial biotechnology and whole cell biocatalysis. *Trends in Biotechnology*, 33(4), 237–246. <https://doi.org/10.1016/j.tibtech.2015.02.001>
- Kogan, G., & Kocher, A. (2007). Role of yeast cell wall polysaccharides in pig nutrition and health protection. *Livestock Science*, 109(1–3), 161–165. <https://doi.org/10.1016/j.livsci.2007.01.134>
- Kondo, K., Saito, T., Kajiwara, S., Takagi, M., & Misawa, N. (1995). A transformation system for the yeast *Candida utilis*: Use of a modified endogenous ribosomal protein gene as a drug-resistant marker and ribosomal DNA as an integration target for vector DNA. *Journal of Bacteriology*, 177(24), 7171–7177. <https://doi.org/10.1128/jb.177.24.7171-7177.1995>
- Krahulec, J., Lišková, V., Boňková, H., Lichvariková, A., Šafranek, M., & Turňa, J. (2020). The ploidy determination of the biotechnologically important yeast *Candida utilis*. *Journal of Applied Genetics*, 61(2), 275–286. <https://doi.org/10.1007/s13353-020-00544-w>

- Kunigo, M., Buerth, C., & Ernst, J. F. (2015). Secreted xylanase XynA mediates utilization of xylan as sole carbon source in *Candida utilis*. *Applied Microbiology and Biotechnology*, *99*(19), 8055–8064. <https://doi.org/10.1007/s00253-015-6703-1>
- Kunigo, M., Buerth, C., Tielker, D., & Ernst, J. F. (2013). Heterologous protein secretion by *Candida utilis*. *Applied Microbiology and Biotechnology*, *97*(16), 7357–7368. <https://doi.org/10.1007/s00253-013-4890-1>
- Kurtzman, C. P., Johnson, C. J., & Smiley, M. J. (1979). Determination of conspecificity of *Candida utilis* and *Hansenula jadinii* through DNA reassociation. *Mycologia*, *11*(4), 844–847. <https://doi.org/10.1080/00275514.1979.12021081>
- Kurtzman, Cletus P., Robnett, C. J., & Basehoar-Powers, E. (2008). Phylogenetic relationships among species of *Pichia*, *Issatchenkia* and *Williopsis* determined from multigene sequence analysis, and the proposal of *Barnettozyma* gen. nov., *Lindnera* gen. nov. and *Wickerhamomyces* gen. nov. *FEMS Yeast Research*, *8*(6), 939–954. <https://doi.org/10.1111/j.1567-1364.2008.00419.x>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. (2012). Exploring single-sample snp and indel calling with whole-genome de novo assembly. *Bioinformatics*, *28*(14), 1838–1844. <https://doi.org/10.1093/bioinformatics/bts280>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *00*(00), 1–3. <http://arxiv.org/abs/1303.3997>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Libkind, D., Peris, D., Cubillos, F. A., Steenwyk, J. L., Oplente, D. A., Langdon, Q. K., Rokas, A., & Hittinger, C. T. (2021). Into the wild: New yeast genomes from natural environments and new tools for their analysis. *FEMS Yeast Research*, *20*(2), 1–15. <https://doi.org/10.1093/FEMSYR/FOAA008>

- Lööke, M., Kristjuhan, K., & Kristjuhan, A. (2011). Extraction of genomic DNA from yeasts for PCR-based applications. *BioTechniques*, *50*(5), 325–328. <https://doi.org/10.2144/000113672>
- Margarido, G. R. A., & Heckerman, D. (2015). ConPADE: genome assembly ploidy estimation from next-generation sequencing data. *PLoS Computational Biology*, *11*(4), 1–25. <https://doi.org/10.1371/journal.pcbi.1004229>
- Martínez, E. A., dos Santos, J. F., Araujo, G. S., de Souza, S. M. A., de Cássia Lacerda Brambilla Rodrigu, R., & Canettieri, E. V. (2018). Production of Single Cell Protein (SCP) from Vinasse. 215–238. https://doi.org/10.1007/978-3-319-90379-8_10
- Meyer, H. P. (2011). Sustainability and biotechnology. *Organic Process Research and Development*, *15*(1), 180–188. <https://doi.org/10.1021/op100206p>
- Moftah, O. A. S., Grbavčić, S., Žuža, M., Luković, N., Bezbradica, D., & Knežević-Jugović, Z. (2012). Adding value to the oil cake as a waste from oil processing industry: Production of lipase and protease by *Candida utilis* in solid state fermentation. *Applied Biochemistry and Biotechnology*, *166*(2), 348–364. <https://doi.org/10.1007/s12010-011-9429-2>
- Morata, A., Loira, I., Tesfaye, W., Bañuelos, M. A., González, C., & Suárez Lepe, J. A. (2018). Lachancea thermotolerans applications in wine technology. *Fermentation*, *4*(3). <https://doi.org/10.3390/fermentation4030053>
- Murali, D., & Dhandayuthapani, K. (2019). Optimization of fermentation of *Ulva* sp. hydrolysate by novel yeast *Cyberlindnera jadinii* MMS7 for enhancement of polyphenol content and antioxidant activity. *9*, 1–7.
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, *32*(11), 1749–1751. <https://doi.org/10.1093/bioinformatics/btw044>
- Parkin, M. J., & Ross, I. S. (1986). The specific uptake of manganese in the yeast *Candida utilis*. *Journal of General Microbiology*, *132*(8), 2155–2160. <https://doi.org/10.1099/00221287-132-8-2155>
- Qiu, Y., Hamernick, S., Ortiz, J. B., & Watkins, E. (2020). DNA content and ploidy estimation of *Festuca ovina* accessions by flow cytometry. *Crop Science*, *60*(5), 2757–2767. <https://doi.org/10.1002/csc2.20229>
- Ramirez-Gonzalez, R. H., Bonnal, R., Caccamo, M., & MacLean, D. (2012). Bio-samtools: Ruby bindings for SAMtools, a library for accessing BAM files containing high-throughput sequence alignments. *Source Code for Biology and Medicine*, *7*, 1–6. <https://doi.org/10.1186/1751-0473-7-6>

- Rao, V. S., Das, S. K., Rao, V. J., & Srinubabu, G. (2008). Recent developments in life sciences research: Role of bioinformatics. *African Journal of Biotechnology*, 7(5), 495–503.
- Ribas, D., Sá-Pessoa, J., Soares-Silva, I., Paiva, S., Nygård, Y., Ruohonen, L., Penttilä, M., & Casal, M. (2017). Yeast as a tool to express sugar acid transporters with biotechnological interest. *FEMS Yeast Research*, 17(2), 1–12. <https://doi.org/10.1093/femsyr/fox005>
- Ribeiro, B. G., de Veras, B. O., dos Santos Aguiar, J., Medeiros Campos Guerra, J., & Sarubbo, L. A. (2020). Biosurfactant produced by *Candida utilis* UFPEDA1009 with potential application in cookie formulation. *Electronic Journal of Biotechnology*, 46, 14–21. <https://doi.org/10.1016/j.ejbt.2020.05.001>
- Rodríguez, B., Valdivié, M., Lezcano, P., & Herrera, M. (2013). Evaluation of torula yeast (*Candida utilis*) grown on distillery vinasse for broilers. *Cuban Journal of Agricultural Science*, 47(2), 183–188.
- ROUSH, A. H., QUESTIAUX, L. M., & DOMNAS, A. J. (1959). The active transport and metabolism of purines in the yeast, *Candida utilis*. *Journal of Cellular and Comparative Physiology*, 54, 275–286. <https://doi.org/10.1002/jcp.1030540310>
- Rupp, O., Brinkrolf, K., Buerth, C., Kunigo, M., Schneider, J., Jaenicke, S., Goesmann, A., Pühler, A., Jaeger, K. E., & Ernst, J. F. (2015). The structure of the *Cyberlindnera jadinii* genome and its relation to *Candida utilis* analyzed by the occurrence of single nucleotide polymorphisms. *Journal of Biotechnology*, 211, 20–30. <https://doi.org/10.1016/j.jbiotec.2015.06.423>
- Santiago, C., Rito, T., Vieira, D., Fernandes, T., Pais, C., Sousa, M. J., Soares, P., & Franco-Duarte, R. (2021). Improvement of *Torulasporea delbrueckii* genome annotation: Towards the exploitation of genomic features of a biotechnologically relevant yeast. *Journal of Fungi*, 7(4). <https://doi.org/10.3390/jof7040287>
- Sauer, M., Porro, D., Mattanovich, D., & Branduardi, P. (2008). Microbial production of organic acids: expanding the markets. *Trends in Biotechnology*, 26(2), 100–108. <https://doi.org/10.1016/j.tibtech.2007.11.006>
- Soares-Silva, I., Ribas, D., Sousa-Silva, M., Azevedo-Silva, J., Rendulić, T., & Casal, M. (2020). Membrane transporters in the bioproduction of organic acids: State of the art and future perspectives for industrial applications. *FEMS Microbiology Letters*, 367(15), 1–15. <https://doi.org/10.1093/femsle/fnaa118>
- Soares-Silva, Isabel, Ribas, D., Foskolou, I. P., Barata, B., Bessa, D., Paiva, S., Queirós, O., & Casal, M. (2015). The *Debaryomyces hansenii* carboxylate transporters jen1 homologues are functional in

- Saccharomyces cerevisiae*. *FEMS Yeast Research*, 15(8), 1–10.
<https://doi.org/10.1093/femsyr/fov094>
- Sousa-Silva, M., Soares, P., Alves, J., Vieira, D., Casal, M., & Soares-Silva, I. (2022). Uncovering Novel Plasma Membrane Carboxylate Transporters in the Yeast *Cyberlindnera jadinii*. *Journal of Fungi*, 8(1). <https://doi.org/10.3390/jof8010051>
- Sousa-Silva, M., Vieira, D., Soares, P., Casal, M., & Soares-Silva, I. (2021). Expanding the knowledge on the skillful yeast *Cyberlindnera jadinii*. *Journal of Fungi*, 7(1), 1–19.
<https://doi.org/10.3390/jof7010036>
- Tahir, H. A. S., Gu, Q., Wu, H., Raza, W., Safdar, A., Huang, Z., Rajer, F. U., & Gao, X. (2017). Effect of volatile compounds produced by *Ralstonia solanacearum* on plant growth promoting and systemic resistance inducing potential of *Bacillus* volatiles. *BMC Plant Biology*, 17(1), 1–16.
<https://doi.org/10.1186/s12870-017-1083-6>
- Tan, Y. C., Kumar, A. U., Wong, Y. P., & Ling, A. P. K. (2022). Bioinformatics approaches and applications in plant biotechnology. *Journal of Genetic Engineering and Biotechnology*, 20(1).
<https://doi.org/10.1186/s43141-022-00394-5>
- Watanasrisin, W., Iwatani, S., Oura, T., Tomita, Y., Ikushima, S., Chindamporn, A., Niimi, M., Niimi, K., Lamping, E., Cannon, R. D., & Kajiwarra, S. (2016). Identification and characterization of *Candida utilis* multidrug efflux transporter CuCdr1p. *FEMS Yeast Research*, 16(4), 1–12.
<https://doi.org/10.1093/femsyr/fow042>
- Weib, C. L., Pais, M., Cano, L. M., Kamoun, S., & Burbano, H. A. (2018). nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*, 19(1), 1–8.
<https://doi.org/10.1186/s12859-018-2128-z>
- Wright, B., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., & Grueber, C. E. (2019). From reference genomes to population genomics: Comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics*, 20(1), 1–10.
<https://doi.org/10.1186/s12864-019-5806-y>
- Yamada, Y., Matsuda, M., & Mikata, K. (1995). The Phylogenetic Relationships of *Pichia jadinii*, Formerly Classified in the Genus *Hansenula*, and Related Species Based on the Partial Sequences of 18S and 26S Ribosomal RNAs (*Saccharomycetaceae*). *Bioscience, Biotechnology, and Biochemistry*, 59(3), 518–520. <https://doi.org/10.1271/bbb.59.518>
- Yoshimura, D., Kajitani, R., Gotoh, Y., Katahira, K., Okuno, M., Ogura, Y., Hayashi, T., & Itoh, T. (2019).

Evaluation of SNP calling methods for closely related bacterial isolates and a novel high-accuracy pipeline: BactSNP. *Microbial Genomics*, 5(5), 1–8. <https://doi.org/10.1099/mgen.0.000261>

Zhang, S. Y., & Liu, S. L. (2013). Bioinformatics. *Brenner's Encyclopedia of Genetics: Second Edition*, 1, 338–340. <https://doi.org/10.1016/B978-0-12-374984-0.00155-8>

7. ANEXOS

Tabela A1 - Reações testadas para amplificação de DNAg com uso de diferentes enzimas.

Reação 1	Quantidade (µl)	Reação 2	Quantidade (µl)
Taq II DNA polimerase	0.25	Accuzyme (2x concentrada)	12.5
Tampão (10x concentrado)	2.5	-	-
MgCl ₂ (50 mM)	1.25	-	-
dNTPs (100 mM)	1	-	-
DNAg	1	DNAg	1
H ₂ O U.P.	17	H ₂ O U.P.	10.5
<i>Primer forward</i> (10 µM)	1	<i>Primer forward</i> (10 µM)	0.5
<i>Primer reverse</i> (10 µM)	1	<i>Primer reverse</i> (10 µM)	0.5
Total	25	Total	25

Tabela A2 – Reação de PCR de colónias.

Componente	Quantidade (µl)
Taq II DNA polimerase	1
Tampão (10x concentrado)	5
MgCl ₂ (50 mM)	2.5
dNTPs (100 mM)	1
DNA (várias colónias)	1
H ₂ O U.P.	31.5
<i>Primer GAP forward</i> (5 µM) ^a	4
<i>Primer CYC reverse</i> (5 µM) ^a	4
Total	50

^a - primers fornecidos pelo laboratório

Tabela A3 - Digestão do DNAp p416GPD utilizando as enzimas SmaI e a Hind III.

Componente	Quantidade (μ l)
DNAp	2
Enzima SmaI (10 U/ μ L)	0.5
Enzima Hind III (10 U/ μ L)	2
Tampão <i>FastDigest</i>	2
H ₂ O	12.5
Fosfatase Alcalina	1
Total	20