UNIVERSIDADE DE SOROCABA PRÓ-REITORIA DE GRADUAÇÃO E ASSUNTOS ESTUDANTIS CURSO DE ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA

Matheus Henrique Galvão

PROSPECÇÃO EM BANCOS DE DADOS PÚBLICOS, CLONAGEM E PRODUÇÃO DE PROTEÍNAS ANFIPÁTICAS RECOMBINANTES EM *ASPERGILLUS ORYZAE*

Sorocaba/SP

2025

Matheus Henrique Galvão

PROSPECÇÃO EM BANCOS DE DADOS PÚBLICOS, CLONAGEM E PRODUÇÃO DE PROTEÍNAS ANFIPÁTICAS RECOMBINANTES EM *A. ORYZAE*

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do Diploma de Graduação em Engenharia de Bioprocessos e Biotecnologia, da Universidade de Sorocaba.

Orientador: Dr. Fábio Marcio Squina
Coorientadora: Dra. Thais Jardim Oliveira

Sorocaba/SP
2025

Matheus Henrique Galvão

PROSPECÇÃO EM BANCOS DE DADOS PÚBLICOS, CLONAGEM E PRODUÇÃO DE PROTEÍNAS ANFIPÁTICAS RECOMBINANTES EM *ASPERGILLUS ORYZAE*

Trabalho de Conclusão de Curso apresentado como exigência parcial para obtenção do Diploma de Graduação em Engenharia de Bioprocessos e Biotecnologia, da Universidade de Sorocaba.

Aprovado em: 05/06/2025

BANCA EXAMINADORA:

_____

Prof.(a) Dr. Fábio Marcio Squina
Universidade de Sorocaba

_____

Prof.(a) Dra. Thaís Jardim Oliveira
Universidade de Sorocaba

_____

Prof.(a) Dra. Tatiane Pasquoto Stigliani
Universidade de Sorocaba

Dedico este trabalho a todas as pessoas que, direta ou indiretamente, contribuíram para minha trajetória acadêmica e pessoal.

Primeiramente, agradeço a Deus pela dádiva da vida e pela oportunidade de desvendar, por meio da ciência, os mistérios do universo que Ele criou. Manifesto, igualmente, minha gratidão à minha família, que desde o meu nascimento tem dedicado tempo e esforço para me oferecer a base necessária para enfrentar os desafios da vida e alcançar meus objetivos.

Sou profundamente grato aos professores que me acompanharam ao longo de minha formação, desde os primeiros anos na pré-escola até a presente etapa, com destaque especial aos educadores da Universidade de Sorocaba, que enriqueceram meu aprendizado em engenharia de Bioprocessos e Biotecnologia.

Por fim, registro meu agradecimento às pessoas que desempenham papéis muitas vezes indiretos, mas fundamentais, na manutenção de um ambiente propício ao aprendizado. Entre elas, a equipe de higienização da universidade, que nos proporciona espaços acolhedores e organizados, e os motoristas de transporte coletivo e vans que asseguram nossa segurança durante os trajetos diários nas vindas e retornos após de longos dias exaustivos de trabalho, especialmente Gilmar Marcos (Gaúcho), Guilherme Santana e Neto Landi

AGRADECIMENTOS

Expresso minha sincera gratidão ao professor Dr. Fábio Marcio Squina, cujo estímulo e apoio desde nosso primeiro contato abriram as portas para o universo da pesquisa científica. Sua dedicação em compartilhar tempo, conhecimento e recursos foi imprescindível para que eu construísse uma base sólida na área, transformando conceitos teóricos aparentemente distantes em experiências práticas concretas e acessíveis.

Ressalto também a contribuição da FAPESP — Fundação de Amparo à Pesquisa do Estado de São Paulo — que, por dois anos consecutivos, 2024 e 2025, concedeu-me bolsa de iniciação científica para o desenvolvimento de projetos de pesquisa. Sou profundamente grato à Fundação pelo papel essencial desempenhado nesta jornada de desenvolvimento acadêmico.

Agradeço ainda às equipes do Laboratório de Ciências Moleculares (LACIMO) da Universidade de Sorocaba, coordenado pelo professor Dr. Fábio Marcio Squina e do Laboratório de Enzimologia e Biologia Molecular de Microrganismos, IB-UNICAMP (LEBIMO), sob a coordenação do professor Dr. André Damásio, da Universidade de Campinas. Todos sempre se mostraram disponíveis para compartilhar conhecimento e oferecer suporte aos projetos realizados.

RESUMO

As proteínas anfipáticas produzidas por fungos filamentosos, possuem amplo potencial biotecnológico devido à sua capacidade de interação com superfícies hidrofóbicas. Este estudo teve como objetivo produzir hidrofobinas e proteínas de ligação à superfície hidrofóbica recombinantes homólogas em *Aspergillus oryzae*, utilizando abordagens integradas de bioinformática, clonagem molecular e edição genômica via CRISPR-Cas9. Foram prospectadas sequências de hidrofobinas em bancos de dados públicos (UniProt, Pfam, InterPro) e desenvolvidas ferramentas computacionais (SSN, CPHunter) para identificar padrões estruturais, como *cysteine loops*. Quatro candidatas foram selecionadas: RoIA, H2, H3 (classe I) e TUB (potencial classe III). As sequências foram amplificadas por PCR, clonadas em *E. coli* e integradas no genoma de *A. oryzae*, utilizando um cassete de expressão com promotor pglaA e sistema de seleção *uidA/ pyrGΔ*. A expressão das proteínas foi confirmada por SDS-PAGE, com secreção eficiente de RoIA (16 kDa) e TUB (20 kDa). A ferramenta CPHunter demonstrou alta acurácia na identificação de hidrofobinas, revelando novas candidatas estruturais para projetos futuros. Os resultados destacam a viabilidade de *A. oryzae* como biofábrica para produção de hidrofobinas, com aplicações em biotecnologia industrial.

Palavras-chave: Hidrofobinas; *Aspergillus oryzae*; CRISPR-Cas9; Bioinformática; Expressão homóloga.

ABSTRACT

The amphipathic proteins produced by filamentous fungi have significant biotechnological potential due to their ability to interact with hydrophobic surfaces. This study aimed to produce recombinant homologous hydrophobins and hydrophobic surface-binding proteins in *Aspergillus oryzae*, using integrated approaches involving bioinformatics, molecular cloning, and genome editing via CRISPR-Cas9. Hydrophobin sequences were mined from public databases (*UniProt, Pfam, InterPro*), and computational tools (*SSN, CPHunter*) were developed to identify structural patterns such as cysteine loops. Four candidates were selected: *RoIA, H2, H3* (class I), and *TUB* (potential class III). The sequences were amplified by PCR, cloned in *E. coli*, and integrated into the genome of *A. oryzae* using an expression cassette with the *pglaA* promoter and selection system *uidA/ pyrGΔ*. Protein expression was confirmed by SDS-PAGE, with efficient secretion of *RoIA* (16 kDa) and *TUB* (20 kDa). The *CPHunter* tool demonstrated high accuracy in hydrophobin identification, revealing new structural candidates for future projects. The results highlight the feasibility of *A. oryzae* as a biofactory for hydrophobin production, with applications in industrial biotechnology.

Keywords: Hydrophobins; *Aspergillus oryzae*; CRISPR-Cas9; Bioinformatics; Homologous expression.

LISTA DE SIGLAS E ABREVIATURAS

CRISPR-Cas9	Ferramenta de edição genética
SSN	Redes de Similaridade de Sequências
PFAM	Banco de dados de famílias de proteínas
INTERPRO	Banco de dados de famílias de proteínas e domínios funcionais.
CPHunter	Cysteine Pattern Hunter
PCR	Reação em Cadeia da Polimerase
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
UV	Ultravioleta
gRNA	RNA guia usado em CRISPR
PBS	Solução Tampão Fosfato Salina
HIS-tag	Cauda de histidina
TAE	Tampão utilizado em eletroforese
CTAB	Brometo de Cetiltrimetilamônio
EDTA	Ácido Etilenodiamino Tetraacético
SDS	Dodecil Sulfato de Sódio, detergente
PVP	Polivinilpirrolidona
RNase A	Enzima que degrada RNA.
X-Gluc	5-Bromo-4-chloro-3-indolil-β-D-glucuronídeo
EggNog	Ferramenta de predição funcional de proteínas
COSI	Sítio Comum de Integração de Genes Sintéticos

SUMÁRIO

1.INTRODUÇÃO E REVISÃO....9
1.1 *Aspergillus* como biofábrica fúngica....9
1.2 Proteínas anfipáticas em fungos 10
1.3 Prospecção de novas proteínas e ferramentas de bioinformática 12
1.4 Proteínas candidatas 13
1.5 Mecanismos de modificação genética em *A. oryzae* ORY7....14
2. OBJETIVOS....16
2.1 Objetivo geral 16
2.2 Objetivos específicos....16
3 MATERIAIS E MÉTODOS 17
3.1 Prospecção de proteínas anfipáticas com o uso de bioinformática 17
3.2 Clonagem de sequências: 20
3.3 Integração gênica em *A.oryzae*....24
4. RESULTADOS....26
4.1 Prospecção de proteínas anfipáticas com o uso de bioinformática 26
4.2 Clonagem de sequências 30
4.3 Integração gênica em *A.oryzae*....32
5.DISCUSSÃO....35
5.1 Bioinformática 35
5.2 Engenharia genética 36
5.3 Evolução das técnicas de transformação em *Aspergillus* 36
5.4 Impacto do sistema de promotores na expressão....37
5.5 Comparação com produção de anticorpos em *A.Oryzae*....37
5.6 Limitações e perspectivas futuras....37
6.CONSIDERAÇÕES FINAIS....39
REFERÊNCIAS: 40
ANEXO A....43
ANEXO B....46

1.INTRODUÇÃO E REVISÃO

1.1 *Aspergillus* como biofábrica fúngica

A produção de proteínas recombinantes tem se mostrado crucial para diversas aplicações na indústria farmacêutica, de alimentos e de biocombustíveis. Tradicionalmente, *Escherichia coli* é o sistema de expressão mais empregado em razão de sua elevada taxa de crescimento, ferramentas moleculares avançadas e capacidade de atingir altas densidades celulares em meios de cultivo econômicos (HATTAB *et al.*, 2021). Apesar disso, certas limitações como a formação de corpos de inclusão, a falta de modificações pós-traducionais e a contaminação por endotoxinas motivam a busca por alternativas eucarióticas, nas quais se destacam os fungos filamentosos do gênero *Aspergillus*, em especial *Aspergillus oryzae* (XU *et al.*, 2020).

O *A.oryzae* possui uma longa história de uso na fermentação de alimentos tradicionais, como molho de soja e saquê, o que se correlaciona ao seu reconhecimento como *Generally Recognized as Safe* (GRAS) pela FDA e ao respaldo do (Organização das Nações Unidas para a Alimentação e Agricultura/ Organização Mundial da Saúde) FAO/WHO, evidenciando sua segurança para produção industrial de enzimas e proteínas recombinantes (JIN *et al.*, 2021; ScienceDirect Topics, 2025). Este organismo exibe elevada capacidade de secreção de proteínas diretamente no meio de cultura, o que simplifica processos de purificação (JIN *et al.*, 2021; XU *et al.*, 2020) bem como eficiente aparato de modificações pós-translacionais, glicosilação e dobramento proteico, indispensáveis para a atividade de proteínas de origem eucariótica (JIN *et al.*, 2021; XU *et al.*, 2020).

Além disso, *A. oryzae* apresenta notável tolerância a condições industriais, crescendo em meios de baixo custo e suportando altas densidades celulares (XU *et al.*, 2020), e, de modo geral, fungos filamentosos demonstram versatilidade metabólica e robustez como fábricas celulares para expressão heteróloga (PUNT *et al.*, 2002). Entretanto, a manipulação genética de *A. oryzae* ainda é desafiadora devido ao seu ciclo de vida assexuado e a estrutura complexa da parede celular, além de elevada atividade de proteases endógenas que podem degradar proteínas heterólogas problema contornado através da construção de cepas deficientes em múltiplas proteases (JIN et al., 2021; FLEIßNER & DERSCH, 2010). Adicionalmente, a morfologia filamentosa pode induzir elevada viscosidade e agregação hifal,

comprometendo a dinâmica de mistura em biorreatores; estudos recentes demonstraram que a deleção de componentes da parede celular (α-1,3-glucanos e galactosaminogalactano) reduz a viscosidade do cultivo e aumenta a produção de proteínas recombinantes (MIYAZAWA *et al.*, 2021).

Em contrapartida, *E.coli* se sobressai pela taxa de geração de cerca de 20 min sob condições otimizadas e pelo vasto arsenal de ferramentas de engenharia genética, que permitem controle preciso de promotores e número de cópias de plasmídeos (HATTAB *et al.*, 2021). Esse sistema pode acumular proteínas recombinantes em até 80 % do peso seco celular, resultando em elevados rendimentos de expressão bruta (DEMAIN & VAISHNAV, 2009). Todavia, a expressão intracelular e a formação de corpos de inclusão exigem etapas adicionais de lise, solubilização e refoldagem; além disso, *E. coli* não executa modificações pós-traducionais (por exemplo, glicosilação) e produz lipopolissacarídeos que requerem purificação rigorosa para aplicações farmacêuticas (U.S. Food and Drug Administration, 2019).

A escolha entre microrganismos como sistema hospedeiro para produção de proteínas recombinantes depende das características bioquímicas da proteína-alvo e das exigências do processo produtivo. Para proteínas simples e não glicosiladas, sem necessidade de secreção, *E. coli,* por exemplo, oferece agilidade e baixo custo operacional (XU *et al.*, 2020).

Por outro lado, para proteínas complexas que demandam modificações pós-traducionais e secreção direta no meio de cultura, *A. oryzae* representa plataforma preferencial, apesar dos desafios de engenharia genética e controle de proteases endógenas (JIN et al., 2021; FLEIßNER & DERSCH, 2010). A contínua evolução de ferramentas de edição genômica e estratégias de engenharia metabólica prometem ampliar ainda mais o potencial de *A. oryzae* como sistema de expressão de proteínas recombinantes em escala industrial.

1.2 Proteínas anfipáticas em fungos

As hidrofobinas são proteínas anfipáticas produzidas por fungos filamentosos, amplamente reconhecidas por seu potencial biotecnológico. Compostas, em sua maioria, por 100 a 150 aminoácidos, essas proteínas desempenham um papel essencial em diversas fases do ciclo de vida fúngico, especialmente na formação de

estruturas aéreas, como hifas, esporos e estruturas reprodutivas (TURGUT & ÖRTÜCÜ, 2023a). As hidrofobinas são classificadas, predominantemente, em duas classes principais: Classe I e Classe II, com a possibilidade de uma Classe III, cuja existência ainda é objeto de debate na literatura (OHTAKI et al., 2006; TURGUT & ÖRTÜCÜ, 2023b).

Os "laços" de cisteína (*Cysteine Loops*) são elementos estruturais cruciais para a estabilidade e funcionalidade das hidrofobinas. A formação de ligações dissulfeto estabiliza a conformação tridimensional dessas proteínas, permitindo a criação de estruturas auto-organizadas, como os *"Rodlets"* característicos das hidrofobinas de Classe I (BALL et al., 2020). Essas ligações também promovem a interação das hidrofobinas com superfícies hidrofóbicas, favorecendo a adesão e a formação de camadas protetoras, essenciais para a sobrevivência dos fungos em diferentes ambientes (TERAUCHI et al., 2020).

As hidrofobinas de Classe I distinguem-se por sua capacidade de formar camadas hidrofóbicas estáveis na superfície de materiais (TAKAHASHI et al., 2005). Essas proteínas são solúveis em ácidos fortes, como ácido trifluoroacético (TFA) e ácido fórmico, e mantêm sua solubilidade em dodecil sulfato de sódio (SDS) a altas temperaturas (LUGONES et al., 1998). Dentro dessa classe, há duas subcategorias: Classe IA, presente em *Ascomicetos*, como o gênero *Aspergillus*, e Classe IB, encontrada em *Basidiomicetos*. Hidrofobinas de Classe IA em *Aspergillus* formam estruturas de *"Rodlets",* enquanto as de Classe IB podem originar monocamadas ou outras conformações distintas (TURGUT & ÖRTÜCÜ, 2023a).

Por sua vez, as hidrofobinas de Classe II, exemplificadas por HFBI e HFBII de *Trichoderma reesei*, são mais solúveis em detergentes, como SDS a 2% ou etanol a 60% (VEREMAN et al., 2021). Essas proteínas desempenham funções importantes na adesão e na interação fúngica com superfícies hidrofóbicas, influenciando processos de patogenicidade e dispersão de esporos (NAKARI-SETÄLÄ et al., 1996). Diferentemente das hidrofobinas de Classe I, não formam *"Rodlets",* mas sim monocamadas auto-organizadas, sendo amplamente distribuídas entre fungos filamentosos *Ascomicetos*, embora raramente identificadas em *Aspergillus* (TURGUT & ÖRTÜCÜ, 2023b).

A proposta de uma Classe III de hidrofobinas baseia-se em proteínas identificadas em fungos, cuja estrutura e função ainda não estão completamente definidas (BALL et al., 2020). Estudos sugerem que essas hidrofobinas apresentam

isofuncionalidade com as classes estabelecidas, mas sem um padrão estrutural bem caracterizado (OHTAKI et al., 2006). Algumas dessas proteínas formam estruturas semelhantes a *"Rodlets",* indicando uma possível relação com as hidrofobinas de Classe I (TAKAHASHI et al., 2015; Tanaka et al., 2022).

As proteínas do tipo hidrofobina apresentam diversas aplicações industriais e biomédicas. Na indústria têxtil, são utilizadas para modificar a hidrofobicidade de tecidos e funcionalizá-los com agentes antimicrobianos, além de atuarem como retardantes de chama na composição de biomateriais feitos com biomassa fúngica (WÖSTEN & SCHOLTMEIJER, 2015). No setor farmacêutico, facilitam a formulação de fármacos hidrofóbicos, melhorando sua biodisponibilidade (SALVE et al., 2022). Ademais, são empregadas na imobilização de moléculas eletroativas e nanomateriais, otimizando ensaios imunológicos e processos biocatalíticos (WANG et al., 2021; PUSPITASARI et al., 2021).

Já as HsbA (*Hydrophobic Surface Binding Protein A*) são proteínas de superfície essencial em fungos filamentosos, como *Aspergillus fumigatus*. Caracteriza-se por domínios ricos em resíduos hidrofóbicos que permitem interações com superfícies apolares, além de regiões anfipáticas que equilibram afinidade por ambientes aquosos e não polares. Localizada na parede celular ou na interface do fungo com o ambiente, a HsbA facilita a adesão a tecidos hospedeiros, formação de biofilmes e proteção contra estresses, como dessecação e agentes imunes. Sua capacidade de mascarar antígenos superficiais contribui para a evasão do sistema de defesa do hospedeiro, reforçando sua relevância na patogenicidade (LAM et al. 2019)

1.3 Prospecção de novas proteínas e ferramentas de bioinformática

A descoberta de novas hidrofobinas com potencial biotecnológico tem sido impulsionada pelo uso de ferramentas de bioinformática avançadas, como um exemplo dessas ferramentas as *“Sequence Similarity Networks”* (SSNs) (ZALLOT et al., 2019). Essas redes permitem a análise detalhada de similaridades entre sequências de proteínas, possibilitando a identificação de hidrofobinas ainda não anotadas e a predição de suas funções com base em características estruturais compartilhadas, como os padrões conservados de ligações dissulfeto formadas pelos resíduos de cisteína (BERGER & SALLADA, 2019). Além disso, podem servir como

base de informações para análises personalizadas utilizando programação computacional.

1.4 Proteínas candidatas

A RolA, uma hidrofobina de classe I, recebeu especial atenção ao longo do projeto devido ao seu extenso estudo na literatura, o que facilitou sua caracterização e permitiu um melhor planejamento experimental. Essa proteína exibe um comportamento dinâmico em superfícies sólidas, podendo formar filmes altamente organizados ou estruturas bastonetes, dependendo do pH e das características do substrato ao qual se adsorve. Suas potenciais aplicações biotecnológicas estão na degradação de polímeros está relacionada à capacidade de recrutar enzimas como a cutinase CutL1, utilização no encapsulamento de fármacos hidrofóbicos auxiliando na estabilização de bioativos e na sua evasão imunológica e, também, no tratamento de superfícies (TANAKA et al., 2022).

Os métodos de extração e purificação da RolA baseiam-se em protocolos tradicionais de isolamento proteico, incluindo a modificação da solubilidade por meio de manipulação do pH e precipitação seletiva. A propriedade de auto-organização da RolA sob diferentes condições pode ser explorada para sua recuperação de substratos hidrofóbicos, possibilitando a purificação por cromatografia sem perda significativa de atividade biológica (TAKAHASHI et al., 2005). O uso de ácidos na desnaturação temporária desse tipo de proteína, como mencionado anteriormente, é uma estratégia recorrente no seu isolamento, uma vez que permite a dissociação de agregados proteicos e o acesso às suas funções específicas (TERAUCHI et al., 2017).

As proteínas HsbA, produzidas por *A.oryzae*, que é o caso da proteína alvo TUB, desempenham um papel fundamental na fixação e na degradação de superfícies hidrofóbicas, como o polímero biodegradável PBSA (polibutileno succinato-co-adipato). Sua principal função é facilitar a ação no *A.oryzae* das cutinases, enzimas hidrolíticas responsáveis pela degradação de polímeros naturais e até sintéticos. HsbA atua como um mediador, promovendo a adsorção da CutL1 na superfície hidrofóbica e aumentando sua eficiência catalítica (OHTAKI et al., 2006).

Embora compartilhe semelhanças funcionais com hidrofobinas, a TUB não pertence a essa família de proteínas. No entanto, sua capacidade de interagir com superfícies hidrofóbicas e recrutar enzimas para degradação de polímeros reflete um

mecanismo semelhante ao da hidrofobina RolA, que também auxilia a CutL1 na degradação do PBSA. A principal diferença entre essas proteínas está em sua estrutura: enquanto hidrofobinas possuem padrões conservados de ligações dissulfeto, HsbA apresenta um arranjo estrutural distinto, mas com funcionalidade convergente (OHTAKI et al., 2006).

1.5 Mecanismos de modificação genética em *A. oryzae* ORY7

A escolha do sítio de integração foi inspirada na plataforma DIVERSIFY, que utiliza o *"Common Synthetic Gene Integration Site"* (COSI) (JARCZYNSKA et al., 2021). Essa lógica se enquadra bem em sistemas de expressão heteróloga em *A.oryzae* e outras espécies do gênero *Aspergillus*. O COSI é uma sequência sintética comum inserida em loci específicos do genoma, no cromossomo 2, projetada para facilitar a integração precisa de cassetes de expressão por meio de recombinação homóloga, utilizando as sequências flanqueadoras A1 e B1. Esse tipo de abordagem é vantajosa porque padroniza o processo de integração: o mesmo cassete de expressão pode ser inserido em diferentes cepas ou até espécies, sem a necessidade de desenvolver novos vetores por completo específicos para cada organismo. Isso economiza tempo e recursos, além de garantir que o gene de interesse seja inserido em uma região genômica controlada, evitando variações indesejadas na expressão do gene (LI et al., 2023).

Com base nessa mesma lógica aplicada no sistema COSI descrito no artigo da plataforma DIVERSIFY, nosso grupo de pesquisa desenvolveu um cassete de integração que foi aplicado nesse projeto. Esse cassete visa estabelecer uma biofábrica fúngica, para realizar a inserção precisa de genes de interesse (JARCZYNSKA et al., 2021). O cassete de integração foi projetado para operar de forma análoga ao COSI, permitindo uma inserção previsível e eficiente do gene em vários *protospacers*, mediada pelo sistema CRISPR/Cas9.

O gene *PyrG* codifica a enzima orotidina-5'-monofosfato descarboxilase, que é essencial para a biossíntese de uracila e pirimidinas (DU et al., 2014). Nesses sistemas utilizou-se uma cepa de *A.oryzae* ORY7 mutante com uma deleção do gene que codifica a enzima orotidina-5'-monofosfato descarboxilase ***pyrGΔ*** (deficiente em uracila) e com o gene *uidA* que codifica a β-glucuronidase inserido no COSI, dentro da região onde se localiza os protospacers do sistema. Essa mutação cria uma

deficiência que pode ser revertida ao transformar o fungo com um cassete de expressão contendo o gene funcional *PyrG*, que após a transformação pode voltar a funcionar de forma episssomal ou se reintegrar aleatoriamente ao genoma dos transformantes. A restauração da biossíntese de uracila permite que apenas células transformadas cresçam em meio sem uracila e pirimidinas, funcionando como um marcador seletivo eficiente para a identificação de transformantes bem-sucedidos (JARCZYNSKA et al., 2021).

Além disso, o sistema *uidA* serve como um marcador de referência que codifica a enzima β-glucuronidase (GUS). Durante o processo de transformação, o gene *uidA* é utilizado para facilitar a identificação de transformantes. As células que contêm o gene *uidA* produzem uma coloração azul quando cultivadas em meio contendo o substrato X-Gluc, devido à atividade da enzima GUS (JARCZYNSKA et al., 2021). Após a inserção do novo gene de interesse no lugar do *uidA*, as células que foram corretamente transformadas não produzem mais a coloração azul, permitindo a identificação visual de transformantes que perderam a atividade do *uidA* e potencialmente integraram o cassete de expressão naquela posição.

Portanto, o sistema *pyrGΔ* é utilizado para a seleção inicial das células transformadas, enquanto o sistema *uidA* fornece uma forma adicional de seleção visual e validação, permitindo a identificação de transformantes que mais provavelmente contêm o novo gene de interesse. Essa combinação de sistemas de seleção aumenta a eficiência e a precisão do processo de transformação genética (JARCZYNSKA et al., 2021; Li et al., 2023).

2. OBJETIVOS

2.1 Objetivo geral

Produzir proteínas anfipáticas homólogas recombinantes em uma biofábrica fúngica de *A.oryzae.*

2.2 Objetivos específicos

• Prospectar proteínas da literatura e via estudos *in silico.*

• Obter as sequências homólogas por meio de PCR.

• Clonar as sequências em *E.coli.*

• Transformar o *A. oryzae* genomicamente através de CRISPR-Cas9.

• Verificar a expressão e secreção adequada das proteínas recombinantes no meio de cultura.

3 MATERIAIS E MÉTODOS

3.1 Prospecção de proteínas anfipáticas com o uso de bioinformática

Inicialmente, foram escolhidas, como proteínas de referência, com base na literatura e buscas no UniProt, cinco proteínas hidrofobinas, já caracterizadas, para criar pontos de orientação na SSN e nas análises de isofuncionalidades. Essas proteínas foram: SC3 de *Schizophyllum commune* (DE VOCHT et al., 2000) RolA de *A. oryzae* (TAKAHASHI et al., 2005) e VHM3 de *Pleurotus ostreatus* (KULKARNI et al., 2022) (pertencentes a classe I), e HFBI e HFBII de *Hypocrea jecorina* (*Trichoderma reesei*) (VEREMAN et al., 2021) (pertencentes a classe II).

A estratégia adotada para a obtenção do banco inicial de sequências de hidrofobinas envolveu o mapeamento de famílias PFAM e INTERPRO (EL-GEBALI et al., 2019) que apresentam características relacionadas às funções de hidrofobinas e HsbA, em conjunto com o banco de dados do UniProt. Os bancos de dados Pfam e INTERPRO (*Classification of protein families*) contêm informações sobre domínios conservados e sequências de proteínas, sendo utilizados como ponto de partida para a predição funcional de proteínas. As famílias, superfamílias e domínios PFAM relevantes para este estudo incluíram: IPR019778, PF01185, IPR001338, IPR010636, IPR036686, PF06766, IPR021054 e PF12296, que englobam hidrofobinas das classes 1 e 2, além de proteínas com características funcionais relacionadas à classe 3, como as HsbA.

Subsequentemente, essa lista de famílias, superfamílias e domínios PFAM de hidrofobinas foi utilizada para gerar a rede de similaridade de sequência (SSN). Nessas análises, a ferramenta EFI-EST foi utilizada no modo de agrupamento por famílias. Diversas iterações de teste foram realizadas na plataforma EFI-EST (OBERG et al., 2023), variando os parâmetros da SSN, como a opção de cálculo de arestas (*edge calculation option*) e o limiar de pontuação de alinhamento (*alignment score threshold)* (HORNUNG & TERRAPON, 2023).

Considerando que as sequências de hidrofobinas e HsbA são relativamente curtas, os valores desses dois parâmetros foram definidos como 20, correspondendo ao *E-value* e ao corte mínimo do percentual de similaridade, respectivamente, para uma clusterização mais precisa (HORNUNG & TERRAPON, 2023). Um arquivo com extensão .CSV foi gerado pela ferramenta EFI-EST, que foi utilizado na construção da

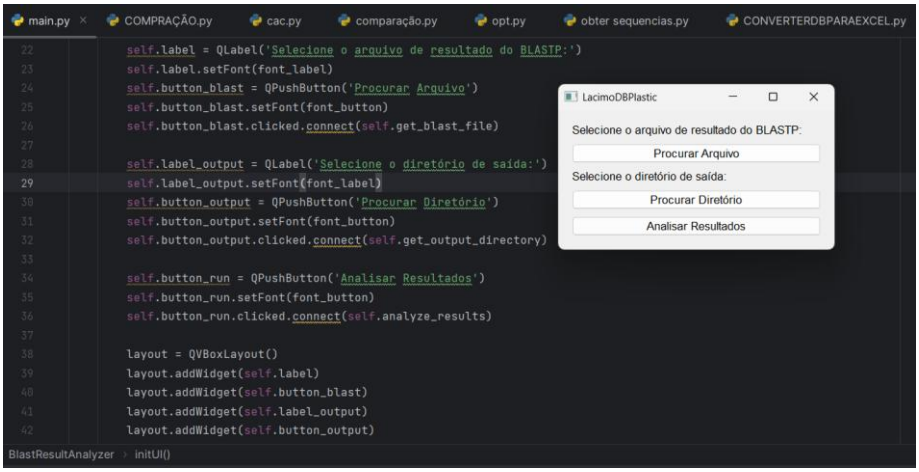
SSN, organizada e visualizada no *Cytoscape* (SHANNON et al., 2003). As análises permitiram distinguir os clusters de maneira coerente com a literatura sobre hidrofobinas, principalmente aquelas tomadas como referência. No total, a rede gerada em nosso estudo contou com 24.686 sequências de hidrofobinas.

Utilizando a ferramenta *EggNog Mapper* (modo WEB via Google) (JENSEN et al., 2008), realizou-se o upload do arquivo (.FASTA) do conjunto de proteínas preditas no genoma de *A.oryzae* (NCBI GCA_000184455.3). A partir da planilha de Excel gerada pela plataforma *EggNog*, foi realizada uma busca manual de Pfams e descritores para hidrofobinas, com o objetivo de identificar sequências codificantes de hidrofobinas nessa linhagem de *Aspergillus*.

Uma estratégia complementar para a prospecção de hidrofobinas foi o desenvolvimento de um algoritmo em linguagem de programação Python, chamado C1, destinado a comparar o conjunto de proteínas preditas no genoma de *A.oryzae* (NCBI GCA_000184455.3) com o banco de dados de 24.686 sequências utilizado nas análises por SSN. Utilizando a opção de exportar tabela no *software Cytoscape* do SSN criado, gerou-se uma nova tabela do Excel abrangente contendo as informações do UniRef90 (*UniProt Reference Clusters)* de todos os nós (sequências de proteínas conectadas por arestas). Esta tabela contém informações detalhadas das proteínas, incluindo IDs e classificação taxonômica (SHANNON et al., 2003). Com o banco de dados contendo 24.686 sequências, foi realizado um *Blastp* local contra todas as proteínas do genoma da RIB40. O resultado foi um arquivo (.txt) contendo dados de identidade de sequência e *E-value* de todas as proteínas de *A.oryzae* comparadas com a base de proteínas anfipáticas descritas no SSN.

Este processo permitiu testar cada uma das sequências anotadas do genoma de *A.oryzae* quanto à similaridade com cada uma das 24.686 sequências de proteínas anfipáticas mapeadas com domínios Pfam e INTERPRO. Para organizar os dados e filtrar os resultados mais relevantes, utilizando Python, selecionaram-se apenas os resultados com *E-value* menor que 10^{-30} e percentual de identidade superior a 50%, criando-se uma tabela com os resultados. Cada rodada de análise teve um tempo médio de execução de 1 hora e 30 minutos (Figura 1).

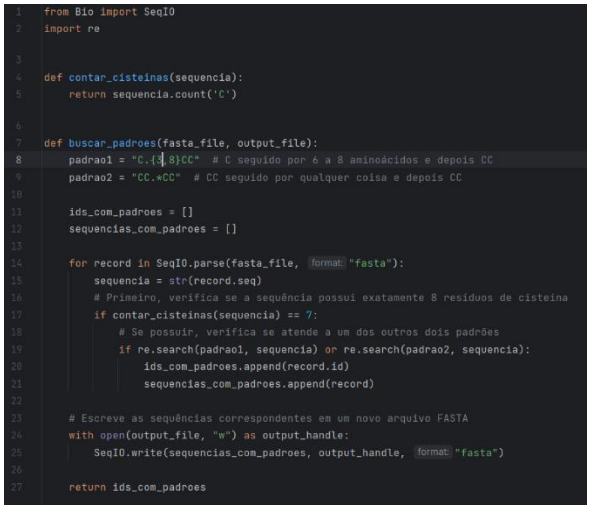
Figura 1 - Código C1 em Python utilizado no desenvolvimento da primeira ferramenta de varredura do genoma das cepas e comparação de cada sequência predita no genoma com o banco de dados local de proteínas criado com *Blast+.*



Fonte - Elaboração própria.

Em um período posterior, desenvolveu-se um algoritmo denominado CPHunter (*CysteinePatternHunter*), capaz de mapear padrões na disposição de cisteínas em genomas, descritos por outros autores como sequências assinatura de hidrofobinas, ou *"cysteine loops"* (TURGUT & ÖRTÜCÜ, 2023b). Desenvolveu-se um código em Python (linguagem regular e Biopython) para buscar sequências que seguissem esse padrão de *"cysteine loops"* no genoma do RIB40. Basicamente, o algoritmo primeiro detecta proteínas no genoma que apresentem aproximadamente 8 cisteínas. Em seguida, o algoritmo seleciona e testa sequências que atendam a dois padrões específicos: uma cisteína seguida por 6 a 8 aminoácidos quaisquer, seguida por duas cisteínas consecutivas (C.{6,8}CC), ou duas cisteínas consecutivas seguidas por qualquer sequência e, em seguida, mais duas cisteínas consecutivas (CC.*CC). Sequências que atendem a esses padrões são armazenadas e exportadas em um novo arquivo FASTA para análise manual ou posterior (Figura 2).

Figura 2 - Código em python utilizado no desenvolvimento da segunda ferramenta CPHunter (*CysteinePatternHunter*) que busca por hidrofobinas dentro das sequencias de proteínas preditas no genoma utilizando linguagem regular baseada nos Loops de Cisteína característicos de proteínas hidrofobinas.



```python
from Bio import SeqIO
import re


def contar_cisteinas(sequencia):
    return sequencia.count('C')


def buscar_padroes(fasta_file, output_file):
    padrao1 = "C.{3,8}CC"  # C seguido por 6 a 8 aminoácidos e depois CC
    padrao2 = "CC.*CC"  # CC seguido por qualquer coisa e depois CC

    ids_com_padroes = []
    sequencias_com_padroes = []

    for record in SeqIO.parse(fasta_file, format: "fasta"):
        sequencia = str(record.seq)
        # Primeiro, verifica se a sequência possui exatamente 8 resíduos de cisteína
        if contar_cisteinas(sequencia) == 7:
            # Se possuir, verifica se atende a um dos outros dois padrões
            if re.search(padrao1, sequencia) or re.search(padrao2, sequencia):
                ids_com_padroes.append(record.id)
                sequencias_com_padroes.append(record)

    # Escreve as sequências correspondentes em um novo arquivo FASTA
    with open(output_file, "w") as output_handle:
        SeqIO.write(sequencias_com_padroes, output_handle, format: "fasta")

    return ids_com_padroes
```

Fonte - Elaboração própria

3.2 Clonagem de sequências:

Devido a necessidade de obtenção de DNA genômico, de alta qualidade, do *A.oryzae* RIB40, inicialmente fez-se a extração do material genético utilizando um protocolo adaptado especificamente para esse tipo de microrganismo. Os materiais necessários para a extração incluíram nitrogênio líquido, D-Sorbitol, Tris-Base (1M), HCl 37%, EDTA (0,5 M), cloreto de sódio (NaCl), brometo de cetil trimetilamônio (CTAB), Sarkosyl 5% (200 mL) ou SDS, polivinilpirrolidona (PVP 1% - 200 mL), proteinase K, acetato de potássio (5M, pH 7,5 - 250 mL), fenol:clorofórmio:álcool isoamílico (25:24:1), acetato de sódio (3M), etanol 96%, etanol 70%, isopropanol, hidróxido de sódio e RNase A (FULTON et al., 1995; MURAKAMI et al., 1991).

Preparação das Soluções: Tampão A: 31,9 g de sorbitol, 50 mL de Tris-HCl 1M (pH 9), 5 mL de EDTA 0,5M (pH 8) e água destilada até completar 500 mL. Tampão B: 100 mL de Tris-HCl 1M (pH 9), 50 mL de EDTA 0,5M, 58,44 g de NaCl, 10 g de CTAB e água destilada até completar 500 mL. As soluções tampão foram preparadas e autoclavadas.

Tampão de Lise: Para cada 10 mL de amostra, utilizou-se 3,75 mL do Tampão A, 3,75 mL do Tampão B, 1,5 mL de Sarkosyl 5%, 1 mL de PVP 1% e 100 µL de proteinase K. O tampão de lise foi preparado imediatamente antes do uso e mantido a 65 °C.

A amostra de micélio foi transferida para um almofariz, moída com nitrogênio líquido e transferida para um tubo Falcon de 50 ml. Então o tampão de lise foi adicionado à amostra e a mistura foi incubada a 65 ºC por 10 minutos. Após a incubação, o acetato de potássio foi adicionado e a solução foi incubada por 15 minutos no gelo. A solução foi então centrifugada por 15 minutos a 4ºC a 5000 rpm e o sobrenadante foi transferido para um novo tubo Falcon, onde foi adicionado Fenol:Clorofórmio:Álcool isoamílico na proporção de 25:24:1 respectivamente. Após a centrifugação por 15 minutos a 4ºC a 5000 rpm, a fase superior foi transferida para um novo tubo e o processo de lavagem com Fenol:Clorofórmio:Álcool foi repetido. (Fulton et al., 1995; Murakami et al., 1991). A RNase A foi adicionada e a solução foi incubada à temperatura ambiente.

Em seguida, o acetato de sódio e o etanol 100% gelado foram adicionados e a solução foi incubada no gelo por 15 minutos. Após a centrifugação, o sobrenadante foi descartado, o pellet foi lavado com etanol 70% e seco à temperatura ambiente. Finalmente, o pellet foi dissolvido em 500 uL de TE e incubado por uma noite (Fulton et al., 1995; Murakami et al., 1991). A amostra de DNA resultante, no dia seguinte, foi então transferida para um tubo de 1,5 ml e uma porção de 1 uL do produto foi aplicado em um gel de agarose 1% com 4 uL de brometo de etídio para verificação da qualidade do produto final obtido.

No mesmo período da extração do DNA genômico, em posse do genoma completo de *A.oryzae* disponível no NCBI GCA_000184455.3, foram identificadas as sequências de interesse selecionadas em análises de bioinformática mencionadas anteriormente (Anexo A). Empregando a ferramenta PRIMER3 (Web), essas sequências foram utilizadas para a construção dos primers para PCR convencional que amplificavam a sequência desconsiderando seu peptídeo sinal natural. Também, produziu-se primers para a geração de fragmentos para clonagem via sistema Gibson usando a ferramenta *NEBuilder*. Então fez-se a aquisição dos primers da tabela 1 com a empresa *Exxtend* (Tabela 1).

Tabela 1 - A tabela traz os conjuntos de primers para PCR utilizados para obtenção das sequências de proteínas alvo em *A.oryzae* e os primers utilizados na preparação das sequências para as clonagens utilizando a técnica de Gibson.

Primers utilizados para isolar sequências		
PRIMERS	**F**	**R**
RolA	5'ATGCAGTTCTCCGTCGCC3'	5'CAGGAGAGAGCCGAGAG3'
TUB	5'ATGCAACGCTCGCGAATTAA3'	5'TACGGAGTAGGGGGCGATTG3'
H2	5'ATGCACAGCACCAACATTTTCA3'	5'GTCGCCCTTCTTGCCACT3'
H3	5'ATGAAGGTTTCTACCTTGTTGACC3'	5'AGCCAAGAAGCCAAGAGTAGT3'
Primers Utilizados na técnica de Gibson		
RolA	5'TGGCAAATGTGATTTCCAAGCTTCCTCCTGCCTCTGGC3'	5'TAGTGGTGGTGATGATGGTGCAGGAGAGAGCCGAGAGC3'
TUB	5'TGGCAAATGTGATTTCCAAGATGCAACGCTCGCGAATTAAC3'	5'TAGTGGTGGTGATGATGGTGTACGGAGTAGGGGGCGATTG3'
H3	5'TGGCAAATGTGATTTCCAAGGCGGATCCAACCGCACAATG3'	5'TAGTGGTGGTGATGATGGTGAGCCAAGAAGCCAAGAGTAGTC3'

Fonte - Elaboração própria

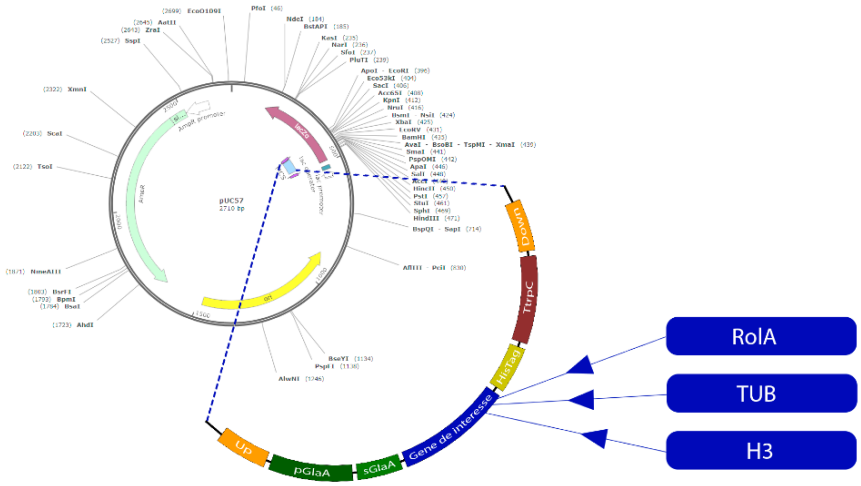
Os produtos de PCR (*polymerase chain reaction*) obtidos diretamente do genoma, referente as três sequências candidatas, foram purificados gel de agarose corado com cristal violeta. Para isso, adicionou-se 50 mg de agarose a 50 ml de tampão TAE e se aqueceu a mistura no micro-ondas por 50 segundos em potência máxima. Em seguida, aguardou-se a temperatura alcançar cerca de 50°C e adicionou-se 15 uL de cristal violeta (P.A) e se agitou levemente. Sequencialmente, a mistura ainda líquida foi vertida na cuba de eletroforese.

Após a solidificação do gel, usaram-se três poços para adicionar, em cada um deles, 25 µL de cada amostra resultante da PCR. Cada amostra foi misturada com 5 uL de *loading* sem corante, H3, TUB e RolA, separadamente. Então, fez-se uma corrida do gel por 30 minutos até se pudesse ver as bandas alvos (olho nu) com coloração azulada. Nesse momento, se fez a excisão precisa da banda do gel usando um bisturi cirúrgico e então a banda foi purificada usando o kit de purificação da *QIAGEN - QIAquick PCR & Gel Cleanup Kit*. A sequência de H2 não pôde ser isolada por PCR, possivelmente devido a uma falha na temperatura de anelamento dos primers.

Em seguida, o plasmídeo pUC57 contendo o cassete de integração de *A. oryzae* (Figura 3), foi linearizado por PCR utilizando os primers (R/pUC57/GLASP: 5’CTTGGAAATCACATTTGCCAAC3’ e F/pUC57/HISTAG 5’ CACCATCATCACCACCACTAAGGATCC 3’). O produto de PCR da linearização do plasmídeo foi purificado, também, pela mesma técnica descrita acima, usando o kit de purificação da *QIAGEN - QIAquick PCR & Gel Cleanup Kit.* O cassete de expressão usado no projeto apresenta um forte promotor glucamilase de *A.oryzae*, uma sequência de peptídeo sinal da glucamilase e o gene de interesse em fusão com uma His-tag (visando futura purificação cromatográfica), e o terminador do gene trpC (do

fungo *A. nidulans*), que é uma sequência finalizadora da transcrição. Este cassete é flanqueado por regiões de homologia (UP e DOWN), que são segmentos específicos de DNA correspondentes a uma região alvo onde ocorre a integração no genoma de *A.oryzae,* via recombinação homóloga (YAO et al., 2024).

Figura 3 - Mapa do vetor pUC57 portando o cassete de integração para *A.oryzae* e as sequências alvo.



Fonte - Elaboração própria

A próxima etapa envolveu a inserção das sequências alvo no cassete de integração do pUC57. Então, foi realizada, novamente, a PCR de cada sequência (H3,RolA e TUB), porém com primers específicos para clonagem pela técnica de Gibson (Tabela 1)(LOPEZ, 2024) . Após a PCR, os produtos foram purificados em gel novamente com o kit *QIAGEN – QIAquick PCR & Gel Cleanup Kit.* Então, tanto o plasmídeo quanto as sequências já preparadas para técnica de Gibson, foram tratados com a enzima DPNI - R0176S da *New England Biotechnology*. Esse tratamento com DPNI tem o objetivo de remover o DNA metilado ou hemimetilado do *template* original utilizado na PCR. Isso evita que ocorra a transformação bacteriana do DNA plasmidial residual do *template*, permitindo maior eficiência no processo de clonagem.

A técnica de Gibson foi realizada com o kit (E5510) *DNA Assembly Master Mix da New England Biotechnology* (LOPEZ, 2024). Então, imediatamente a transformação foi realizada utilizando células termocompetentes de *E. coli* Dh5α. As células foram incubadas com 5 µL do produto da reação de Gibson no gelo por 30

minutos, seguidas de choque térmico a 42ºC por 2 minutos e retorno ao gelo por 4 minutos. Em seguida, 100 µL das células foram transferidos para 900 µL de meio SOC e recuperadas sob agitação a 37ºC por 1 hora. Então, foram feitas placas com 20 mL de meio LB contendo 60 µL de ampicilina (100 mg/mL) e incubadas durante a noite a 37ºC. As colônias transformadas foram isoladas.

3.3 Integração gênica em *A.oryzae*

Em posse dos plasmídeos (pUC57) contendo a sequência das proteínas inseridas no cassete de integração em *A. oryzae*, foram realizadas novos minipreps utilizando *kit PureYield™ Plasmid Miniprep System* – A1223 - Promega. Os minipreps foram tratados com enzima de restrição SWAI - R0604S da *New England Biotechnology*, *overnight*, para linearização. O plasmídeo foi desenvolvido já direcionando o ponto de clivagem da SWAI para gerar como produto o cassete de integração completo com as regiões de homologia para recombinação. A transformação do *A.oryzae* contou com a ajuda e supervisão do aluno de doutorado Gabriel Motta da Universidade Estadual de Campinas (UNICAMP), orientado do Prof. André Damásio.

Para a produção dos protoplastos de *A.oryzae* ORY7- *pyrGΔ*, *ku70Δ*, IS1::A-PgdaA-*uidA*-TtrpC-B, iniciou-se coletando conídios de uma placa contendo meio MMG (1M sacarose, sais de nitrato de *Clutterbucks*, elementos traços, 5 mM uridina, 2 mM uracila, e 2% agar bacteriano). Esses conídios foram transferidos para um frasco de agitação contendo 100 mL de meio YPD (2% glicose, 0,5% extrato de levedura, 2% peptona). A cultura foi incubada a 30°C e agitada a 150 rpm por uma noite. Após a incubação, os micélios germinados foram coletados e lavados com tampão de protoplastação de *Aspergillus* (APB), que contém MgSO4 a 1,1 M e tampão fosfato de sódio a 10 mM, ajustado para pH 5,8 com NaOH 2N. Este tampão estabiliza osmoticamente os protoplastos durante a digestão enzimática.

Em sequência, o micélio foi ressuspendido em solução de APB contendo 40 mg/mL de "VinoTaste PRO" (Novozymes), um complexo enzimático que digere a parede celular fúngica. A suspensão foi homogenizada suavemente e incubada a 30°C por 2-3 horas com agitação a 150 rpm. Após esse período, a mistura foi filtrada através de *Miracloth* para separar os protoplastos, que foram então centrifugados a 3000xg por 12 minutos. O sedimento de protoplastos foi transferido para um novo tubo

e ressuspendido em tampão de transformação de *Aspergillus* (ATB), que contém 1,2 M sorbitol, 50 mM $CaCl_2 \cdot 2H_2O$, 20 mM Tris, e 0,6 M KCl, ajustado para pH 7,2 com HCl 2N. O ATB estabiliza os protoplastos e facilita a absorção do DNA durante a transformação.

Para a transformação em *A.oryzae* ORY7, 50 µL de protoplastos na concentração de 10^9 foram misturados com 10 µL(250 ng/uL) do cassete de integração linearizado com SWAI e 10 µL do vetor de expressão pfc330. O vetor pfc330 contém a sequência que codifica a nuclease Cas9 e o RNA guia (gRNA) para região COSI e devolve o gene *pyrG*. Os melhores resultados foram obtidos misturando os RNA guia para os *protospacers* 1 e 3. A mistura foi realizada em tubos de *Eppendorf* aos quais foram adicionados 150 µL de PCT (50% PEG 8000, 50 mM $CaCl_2$, 20 mM Tris, 0,6 M KCl, pH 7,5), o que promove a fusão entre as membranas celulares e o DNA exógeno, pois o PEG desestabiliza temporariamente a membrana plasmática dos protoplastos, permitindo a entrada do DNA. A solução foi homogenizada invertendo o tubo e incubada por 10 minutos à temperatura ambiente.

Posteriormente, foram adicionados 250 µL de ATB para a mistura, que foi então plaqueada em meio MMG seletivo sem uridina e uracila para a seleção de transformantes. As placas foram incubadas a 30°C por 8 dias. Em seguida as colônias que positivaram para transformação em meio sem uridina e uracila foram inoculadas, por mais 8 dias, em MMG seletivo com X-Gluc. As linhagens transformadas foram armazenadas a -80°C em estoques criogênicos, garantindo sua viabilidade para futuras aplicações e testes de validação.

Em sequência as linhagens transformantes foram cultivadas em meio mínimo MMG (5 ml de clutterbucks, 100 uL de elementos traços, 20uL de uridina e uracila a 55 mg/ml, 1g de glicose e 8ug de proteínas para 94,9 ml de água e pH: 6.5) por 3 dias, sob agitação de 150 rpm, a 30ºC, até haver crescimento da biomassa. Então, a biomassa foi lavada com tampão PBS estéril e transferida para outro meio que contém 5 ml de Clutterbucks , 100 uL de elementos traços, 20uL de uridina e uracila a 55 mg/ml, 8ug de proteínas e 2g de maltose, para 94,9 ml de água e pH: 6.5. O cultivo se dá por mais 72 horas e nesse momento a maltose induz a via do cassete de integração por meio do promotor da Glucamilase *pglaA* e o peptídeo sinal *psglaA* induz a secreção da proteína no meio de cultura. Ao final do período de cultivo, o meio foi filtrado, com Miracloth – Sigma Aldrich, em seguida 0,5 ml foi concentrado para 100

0,1 ml com uma coluna UFC5010 Millipore Filtro para centrífuga Amicon® Ultra, 10 kDa MWCO a 16000 rmp e aplicou-se 20 uL no gel de SDS-page.

4. RESULTADOS

4.1 Prospecção de proteínas anfipáticas com o uso de bioinformática

Durante a avaliação dos resultados gerados comparando o genoma de *A. oryzae* contra o banco com mais de 24 mil sequências prospectadas, usando o código C1, foi possível identificar uma hidrofobina adicional nesse *A. oryzae* (H3- Q2UD05: Tabela 2 – destacada em branco), que não foi detectada pela busca manual no genoma de *A. oryzae* utilizando a ferramenta *EggNog*. Isso porquê, se tratando de uma sequência muito curta, nos parâmetros padrões da *EggNog*, ela não é identificada.

Tabela 2 - Resultado *Output* gerado pela análise feita a partir do código C1. Em cinza escuro pode-se observar as sequências que já haviam sido detectadas na análise manual usando a planilha *EggNog* e foram novamente encontradas pela ferramenta, o que confirma a acurácia do código. Em branco, destaca-se a proteína Q2UD05 que foi nomeada de H3.

Query	UniREf90	Description	Alvos
XP_001821309.3	Q2UGK8	Hydrophobic surface binding protein A OS=A...	
XP_001824881.1	Q873N6	Hydrophobin OS=Aspergillus oryzae OX=506...	RolA
XP_001825165.1	Q2U333	Cell wall galactomannoprotein OS=Aspergill...	
XP_001825181.1	Q2U317	Hydrophobic surface binding protein A-doma...	
XP_001825805.1	Q2U193	Hydrophobic surface binding protein A-doma...	
XP_023089925.1	Q2ULN2	Hydrophobic surface binding protein A OS=A...	TUB
XP_023091183.1	B8NFY5	Hydrophobin OS=Aspergillus flavus (strain ...	
XP_023091382.1	**Q2UD05**	**Hydrophobin OS=Aspergillus oryzae (strai...**	**H3**
XP_023091658.1	Q2UB99	Cell wall galactomannoprotein OS=Aspergill...	
XP_023092217.1	Q2U814	Hydrophobin OS=Aspergillus oryzae (strain ...	H2
XP_023093403.1	Q2U130	Cell wall galactomannoprotein OS=Aspergill...	
XP_023090756.1	B8N9R7	Clathrin heavy chain OS=Aspergillus flavus...	
XP_001819536.1	A0A2P2GWI3	Hydrophobic surface binding protei...	

Fonte - Elaboração própria.

Ao final dessa rodada de prospecção, entre os resultados obtidos pela análise via *EggNog* e a ferramenta construída usando *Blast+*, quatro sequências candidatas foram definidas como alvo do presente projeto: três hidrofobinas clássicas (classe 1), RolA, H2 , H3 e uma proteína hidrofóbica de ligação à superfície, denominada TUB, que é uma potencial classe 3. As ferramentas aplicadas mostraram-se eficazes e podem ser usadas para identificar novas proteínas em outras espécies em projetos futuros. A partir dessas quatro candidatas partiu-se para a amplificação das sequências por PCR.

Já no segundo semestre de 2024, usando outra ferramenta produzida com linguagem regular em Python, avaliamos se poderiam haver hidrofobinas adicionais no genoma de *A. oryzae*. Utilizando-se de linguagem regular, buscando padrões de sequência, *cysteine loops*, relacionados ao número e localização de resíduos de cisteína, apresentou uma série de novas proteínas candidatas a hidrofobinas. As proteínas encontradas tem em sua estrutura resíduos de cisteína que atendem o padrão observado em hidrofobinas, sugerindo sua predição funcional com tal (Tabela 3).

Tabela 3 - A tabela apresenta o resultado da ferramenta *CysteinePatternHunter (CPHunter).* Aparecem na lista todas novas sequências que atenderam o padrão do *cysteine loops* e podem ser potenciais hidrofobinas.

ID	NOME	FUNÇÃO PREDITA (Ref90)	TAMANHO	PFAM	MASSA kDa
Q2TW45		methylaspartate ammonia-lyase	416	PF07476;PF05034;	45,8
Q2U051		Major facilitator superfamily (MFS)	437	PF07690;	47,6
Q2U1J3		Major facilitator superfamily (MFS)	368	PF07690;	39,9
Q2U2B9		Secreted protein	113		11,7
Q2U3W7	RolA	Hydrophobin	151	PF01185;	15,2
Q2U4U0		Phosphatidylinositol N-acetylglucosaminyltransferase	336	PF06432;	37,3
Q2U814	H2	Hydrophobin	116		12,3
Q2UA60		Hydrophobin	111		11,7
Q2UB26		C2H2-type domain-containing protein	146		16,1
Q2UD05	H3	Hydrophobin	108	PF01185;	11,3
Q2UDK2		Hydrophobin	145		14,3
Q2UDR9		Uncharacterized protein	414		48,4
Q2UEU5		Mob1/phocein	463	PF03637;	50,9
Q2UHA4		Enoyl reductase (ER) domain-containing protein	353	PF08240;PF00107;	38,3
Q2UKG9		Major facilitator superfamily (MFS)	355	PF07690;	38,2
Q2UNW5		Probable feruloyl esterase A	281	PF01764;	30,2
Q2UP76		Hydrophobin	130		13,3
Q2UTJ1		serine--tRNA ligase	494	PF02403;PF00587;	56,2

Fonte - Elaboração própria.

Todas proteínas descritas no resultado acima (Tabela 3), seguem o padrão dos *cysteine loops*. Além disso, como se pode observar na tabela 3, três das sequências alvo do projeto foram encontradas e apontadas na coluna “nome”, em vermelho: Q2U3W7 (RolA), Q2U814 (H2), Q2UD05 (H3), ratificando a acurácia da ferramenta CPHunter. A TUB não segue o padrão de laços de cisteína e então não poderia ser detectada.

Além disso, as sequências Q2UA60, Q2UDK2 e Q2UP76 tem sua função descrita no UniRef90 como hidrofobina, mas não apresentam Pfam característico de hidrofobina, impossibilitando detecção para geração de redes de similaridade usando o Pfam como partida. Isso pode ocorrer por conta de que a grande maioria das ferramentas de bioinformática possuírem um corte padrão para desconsiderar sequências muito curtas, o que dificulta a detecção de hidrofobinas, já que na maioria das vezes são sequências bem reduzidas. A ferramenta desenvolvida para procurar por hidrofobinas diretamente em arquivos de proteínas preditas em genomas conseguiu detectar com sucesso novas sequências de domínio de hidrofobinas como descrito na tabela 3. Essa ferramenta tem potencial para ser otimizada e empregada em novos projetos com diferentes gêneros e espécies de organismos.

No SSN, as hidrofobinas de *A. oryzae* foram encontradas em grupos formados por apenas hidrofobinas de classe I. Proteínas como Q2UKG9, Q2UNW5 e Q2UD05, encontradas pela nova ferramenta CPHunter foram adicionadas manualmente na SSN e se conectaram ao grande cluster que inclui as referências SC3 e Vmh3. Já TUB (Q2ULN2) se localizou em um cluster separado, não conectado a clusters de hidrofobinas já que sua estrutura é significativamente diferente, mesmo que se comporte como uma hidrofobina.

No cluster principal das hidrofobinas de classe II (HFBI e HFBII), não foram encontradas hidrofobinas nativas de *A.oryzae*. Clusters significativos de hidrofobinas de classe II foram marcados com um triângulo verde (Figura 2). RolA se agrupou em um cluster denso de hidrofobinas de classe I, indicando alta similaridade de sequência e possível isofuncionalidade entre os nós. Esse cluster inclui diversas espécies e gêneros de *Ascomicetos* e *Basideomicetos*, sugerindo que estruturas semelhantes a RolA são comuns a muitos fungos.

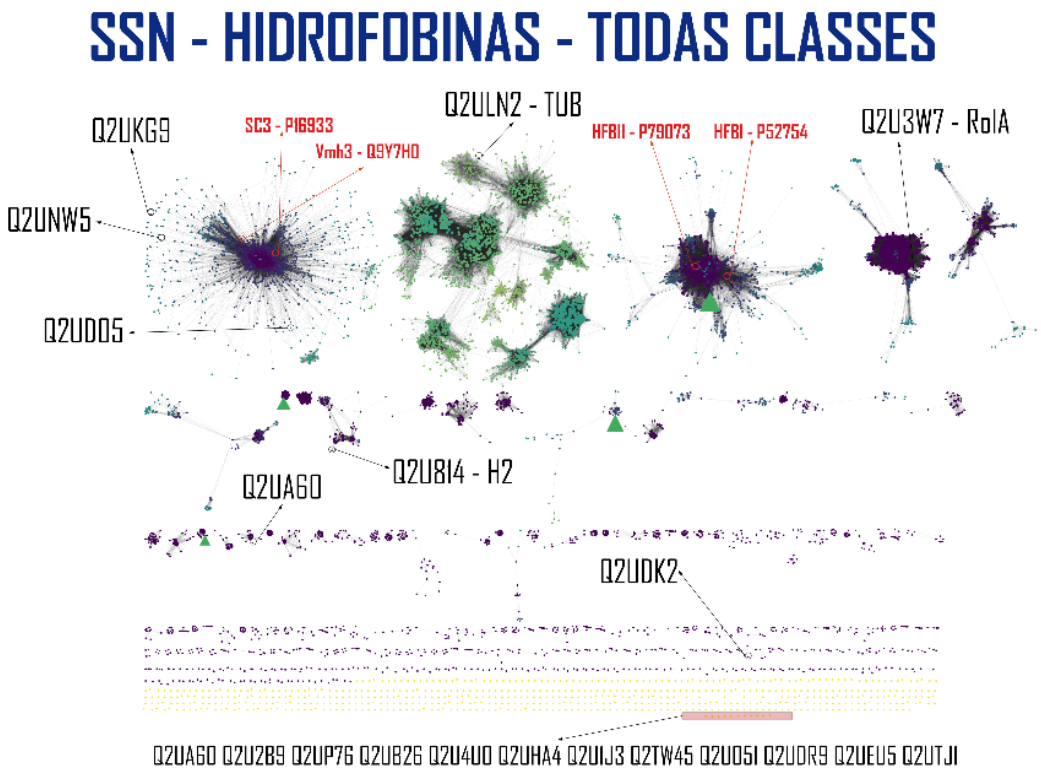
A proteína H2 (Q2U814) foi encontrada em um cluster menor, composto majoritariamente por espécies do gênero *Aspergillus*. Já a Q2UA60 (encontrada pela ferramenta CPHunter) apesentou conexões em cluster pequeno, com proteínas

exclusivas do gênero *Aspergillus*. A Q2UDK2 encontrada pela ferramenta CPHunter, se agrupou apenas com três outros nós do gênero *Aspergillus*.

Outras proteínas (Q2UA60, Q2U2B9, Q2UP76, Q2UB26 encontradas pela ferramenta CPHunter, não formaram arestas com nenhum cluster, apesar de atenderem ao padrão dos laços de cisteínas. Todos nós da porção inferior do SSN, que não se clusterizaram, possuem anotação no UniRef90 como hidrofobina. Isso abre a possibilidade desses nós pertencerem ao grupo de classe III. Isso porque seguem o padrão atual de hidrofobinas, mas por variarem estruturalmente, isso faz com que a ferramenta do EFI-EST e os parâmetros iniciais da SSN acabem tendo dificuldades de organizar esses nós desconectados de forma coerente com o restante dos nós.

Dessa forma, foi possível inserir todas as sequências identificadas por diferentes abordagens e montar uma nova SSN. Algumas proteínas identificadas pela ferramenta: Q2UKG9 e Q2UNW5 se conectaram com o cluster da H3 e estão identificadas no UniRef90 como *"Major facilitator superfamily (MFS)"* e "*Probable feruloyl esterase A"* respectivamente. Isso sugere que essas proteínas apresentam domínios semelhantes às hidrofobinas, fundidos com outros domínios funcionais. Nos casos dessas proteínas foi observado cerca de 35% de identidade de sequência com hidrofobinas de classe I. Coletivamente, podemos concluir o gênero *Aspergillus* apresenta um número significativo de hidrofobinas que puderam ser mapeadas em diversos clusters do conjunto do proteínas analisadas em nosso estudo.

Figura 4 - O SSN final pode ser explorado detalhadamente, cada cluster e nó. Todas hidrofobinas, identificadas até o momento, de *A.oryzae,* foram escritas na coloração preta, as quais os códigos se iniciam com Q2. Em vermelho estão as identificações de todas as proteínas de referências: SC3 de *S. commune* (De Vocht et al., 2000) e VHM3 de *P. ostreatus* (Kulkarni et al., 2022) (pertencentes a classe 1), e HFBI e HFBII de *H. jecorina* (*Trichoderma reesei*) (Vereman et al., 2021) (pertencentes a classe 2). Na barra vermelha inferior estão os nós de hidrofobinas do *A.oryzae* que não se clusterizaram. Os triângulos verdes destacam os cluster contendo hidrofobinas de classe II. (Versão ampliada da imagem nos anexos B)



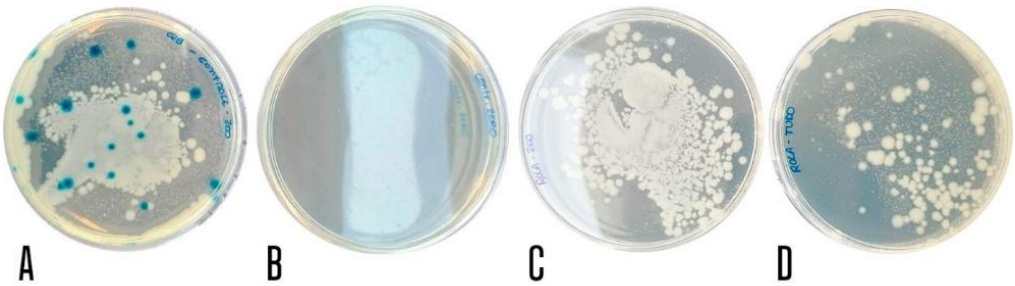
Fonte - Elaboração própria

4.2 Clonagem de sequências

A clonagem das sequências TUB e RolA no vetor pUC57 contendo o cassete de integração foram confirmadas. Contudo, a sequência H3, que foi amplificada com sucesso, não conseguimos confirmar a clonagem no cassete de integração. Diante disso, planeja-se realizar uma nova tentativa clonagem. Como procedimentos para confirmação da clonagem dos fragmentos de proteínas no vetor de integração, foram realizados testes como PCR (não apresentado), digestão com enzima de restrição SWAI (Tabela 4 e Figura 3 - C) e sequenciamento Sanger, apresentado como um blast no NCBI (Figura 4). Estoques criogênicos de culturas de *E. coli* produzindo os

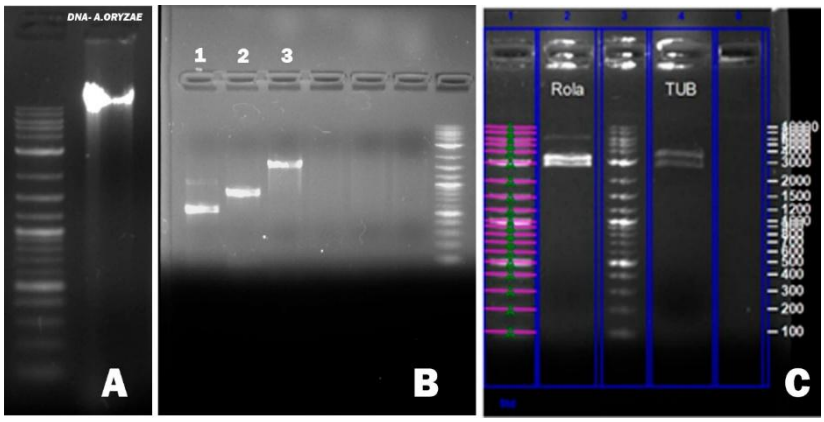
plasmídeos construídos foram feitos em tubos contendo LB com 20% glicerol e armazenados a -80ºC (JANJUA et al., 2014).

Figura 5 - A placas A refere-se ao controle com plasmídeo puc57 que acompanha o kit de Gibson Assembly. A placa B, refere-se ao controle com bactérias não transformadas cultivadas em placas com Ampicilina. A placa C, refere-se à transformação da cepa TUB e a placa D refere-se à **transformação da cepa RolA.**



Fonte - Elaboração própria.

Figura 6 - Na imagem **A**, período de preparação para a clonagem, encontra-se a eletroforese do DNA



genômico extraído de *A.oryzae* RIB40. Na imagem **B**, encontra-se os produtos das PCRs foram realizadas no equipamento T100 Thermal Cycler da marca *Bio Rad*, os kits utilizados foram Q5® High-Fidelity DNA Polymerase e ECRA TUPAM HIFI. A coluna 1 contém o fragmento da proteína RolA - Q2U3W7, a coluna 2 contém o fragmento da proteína TUB - Q2ULN2 e a coluna 3 o fragmento da proteína H3-Q2UD05. Já as colunas de 4 a 6 fazem referência a tentativas de amplificar a proteína H2-Q2U814 em diferentes condições, mas com resultados negativos. Na imagem **C**, destaca-se o gel de eletroforese após a digestão com enzima de restrição SWAI do plasmídeo pUC57 contendo o cassete de integração e os fragmentos de hidrofobinas clonados, na primeira coluna se encontra o marcador, na coluna RolA a digestão plasmídeo. Na terceira coluna, novamente o marcador e em seguida a digestão do plasmídeo contendo o fragmento TUB.

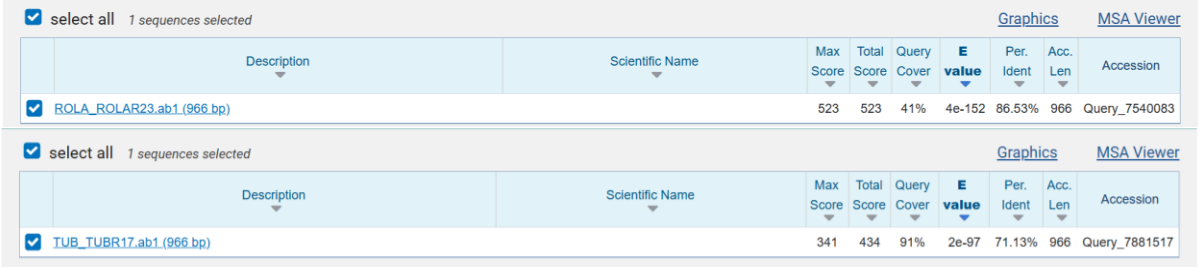
Fonte - Elaboração própria.

Tabela 4 - A tabela destaca o tamanho esperado para o surgimento dos fragmentos após a digestão com a enzima de restrição SWAI, que se pode evidenciar pela Figura 3- C. Os tamanhos esperados dos plasmídeos digeridos que confirmaram a integração da RolA e TUB estão destacados no campo de cor salmão (linhas cassete de integração) da tabela.

FRAGMENTO	Pb	CONTEÚDO
RolA	3182	CASSETE DE INTEGRAÇÃO
TUB	3505	CASSETE DE INTEGRAÇÃO
PUC57	2900	PUC57 + RESÍDEUO CORTE
CASSETE DE INTEGRAÇÃO VAZIO	2670	CASSETE SEM A SEQUÊNCIA ALVO E SEM ALBUMINA
PUC57+CASS+SEQ ROLA	5901	
PUC57+CASS+SEQ TUB	6225	

Fonte - Elaboração própria.

Figura 7 - Confirmação da integração das sequências RolA e TUB (XM_023234847.1) via sequenciamento Sanger e *blast* do resultado do sequenciamento no NCBI. A proteína RolA foi identificada bem como a TUB, em comparação com a sequência do genoma de *A. oryzae*.



Fonte - Elaboração própria.

4.3 Integração gênica em *A.oryzae*

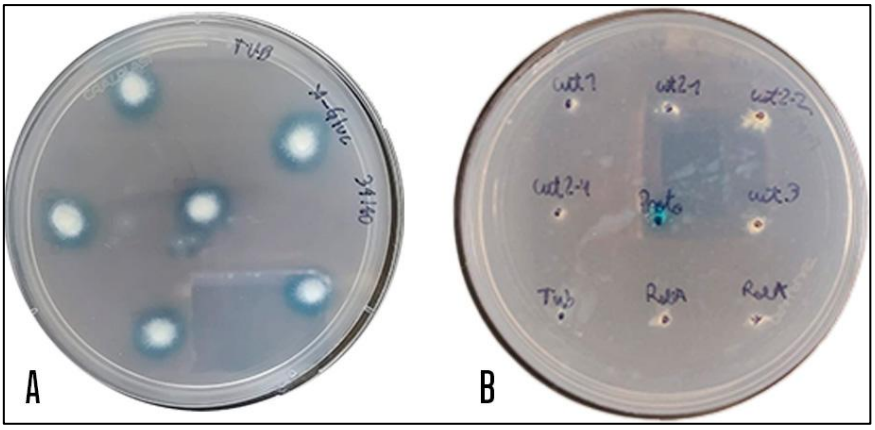
A transformação foi realizada em *A.oryzae* RIB40, utilizando concentrações iniciais de protoplastos de aproximadamente 10^6 células/mL e 2 µg de cassete de integração linearizado com SWAI e 2 µg de vetor de expressão pfc330 com sgRNA para o protospacer 2 apenas. Essa abordagem inicial falhou, como evidenciado pela coloração azul que deriva do sistema de seleção da cepa nas placas contendo X-Gluc (Figura 5 – A)(JARCZYNSKA et al., 2021), indicando que a transformação não foi bem-sucedida. Em uma segunda tentativa, foi decidido aumentar a concentração de protoplastos para 10^9 células/mL e elevar a quantidade de cassete de integração para 10 µg e de vetor de expressão 10 µg e se utilizar sgRNA para os protospacers 1 e 3 juntos. Essa otimização resultou em transformações positivas (Figura 5 - B).

As colônias transformadas foram confirmadas por PCR (Figura 6- B), utilizando primers que flanqueavam as regiões de inserção no cassete de integração

(pGLaA_FWD: 5’GTCGCCTAATGTCTCGTCCGT3’ e TtrpC_REV: 5’GAAGGGCGTACTAGGGTTGCG3’.

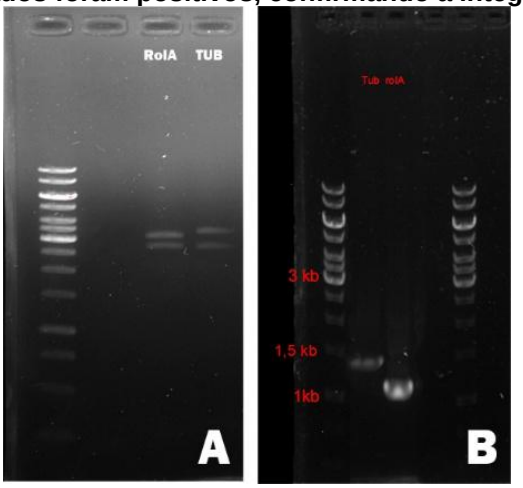
Com isso, conseguimos confirmar a integração gênica das sequências TUB e RolA, gerando cepas fúngicas possivelmente super expressando as proteínas de interesse.

Figura 5 - A imagem “A” refere-se a primeira transforamação em *A.oryzae,* na qual houve falha constatada pela coloração azul. Já na imagem “B”, nota-se apenas o controle na coloração azul no centro, as demais colônias na sua coloração natual indicando o sucesso da integração do cassete (Jarczynska et al., 2021). As demais colônias da placa, com o nome “cut”, referem-se as cutinases do projeto desenvolvido pelo aluno Gabriel Motta, que foram feitas no mesmo dia e serviram também de



Fonte - Elaboração própria.

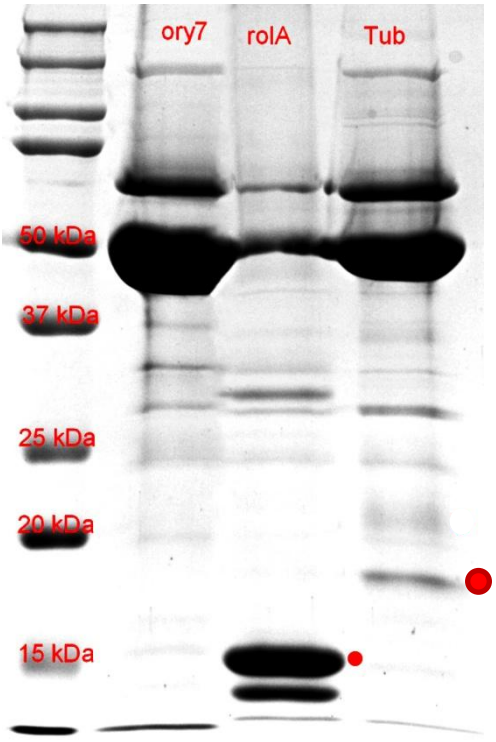
Figura 6 - Na figura 6-A, nas colunas nomeadas RolA e TUB pode-se observar o resultado da digestão com SWAI do plasmídeo pUC57 contendo o cassete de integração antes de transformação em *A.oryzae.* A figura 6-B consiste de resultados de PCR a partir de DNA genômico confirmando que as sequências TUB e RolA foram integradas no genoma de *A. oryzae*. Foram realizadas PCR de confirmação utilizando os primers (pGLaA_FWD: 5’GTCGCCTAATGTCTCGTCCGT3’ e TtrpC_REV: 5’GAAGGGCGTACTAGGGTTGCG3’) que flanqueavam as sequências de interesse inseridas e os **resultados foram positivos, confirmando a integração.**



Fonte - Elaboração própria.

O gel de SDS-page permitiu constatar a secreção das proteínas RolA, como uma banda intensa na região de 16KDa , e a TUB na região de 20KDa como uma banda presente, porém , menos intensa. Esse resultado indica o sucesso da integração do cassete de expressão no genoma do *A.oryzae* e funcionamento adequado da via de secreção adotada no projeto. As condições ótimas de cultivo, visando o aumento do rendimento da produção de proteínas pelas cepas está sendo explorada no momento.

Figura 8 - A figura apresenta o resultado do gel de SDS-page contendo uma alíquota de 0,5 ml concentrada dos meios de indução das cepas transformantes da proteína RolA e TUB. A coluna ORY7 tem-se o controle da cepa não transformada com sequencias de interesse. Na coluna RolA, na altura de 16 KDa tem-se uma banda intensa na região esperada para a proteína RolA. Na coluna TUB, na região de 20 KDa se vê uma banda, não muito intensa, que faz referência a proteína TUB.



Fonte - Elaboração própria.

5.DISCUSSÃO

5.1 Bioinformática

A prospecção de proteínas anfipáticas, como as hidrofobinas, por meio de abordagens bioinformáticas ainda carece de literatura robusta, especialmente quando comparada a outras áreas da biologia computacional. O estudo pioneiro de Littlejohn (LITTLEJOHN et al. 2012) destaca-se como uma das poucas referências metodológicas detalhadas para identificação de hidrofobinas em fungos filamentosos, utilizando estratégias de busca manual em bancos de dados, alinhamentos múltiplos e análises filogenéticas. Seu trabalho revelou a complexidade na anotação dessas proteínas, principalmente devido à baixa homologia além do motivo conservado de cisteínas, e identificou 74 hidrofobinas putativas em espécies de *Aspergillus*, incluindo possíveis formas intermediárias entre as classes I e II. Esse estudo estabeleceu bases críticas para pesquisas subsequentes, mas também evidenciou lacunas, como a necessidade de métodos mais eficientes para lidar com a diversidade estrutural dessas proteínas.

Em comparação, o presente trabalho expandiu as estratégias bioinformáticas ao integrar redes de similaridade de sequência (SSN), domínios Pfam/InterPro e o desenvolvimento de algoritmos customizados em Python (como o C1 e o CPHunter) para detectar padrões de *cysteine loops* em *A.oryzae*. Enquanto Littlejohn e sua equipe focaram em análises manuais e *cross-checking* entre bancos de dados, esta abordagem priorizou ferramentas computacionais automatizadas, como EFI-EST e *blastp* local, permitindo a identificação de outras entidades não detectada por métodos convencionais, além de proteínas candidatas a uma hipotética classe III.

Ambos os estudos concordam na subanotação de hidrofobinas em bancos de dados e na relevância de espécies de *Aspergillus* para aplicações biotecnológicas. Entretanto, este trabalho avançou ao integrar validação experimental via clonagem e transformação genética, além de propor uma ferramenta (CPHunter) para prospecção baseada em padrões estruturais, complementando as limitações das abordagens tradicionais.

Os resultados convergem ao destacar a diversidade de hidrofobinas em *Aspergillus* e a necessidade de métodos refinados para sua identificação. Contudo, as metodologias adotadas são complementares: o estudo de Littlejohn forneceu um

arcabouço teórico e manual, enquanto este trabalho introduziu inovações computacionais e validação prática. Para futuros avanços, a aplicação de inteligência artificial, como redes neurais profundas treinadas em padrões estruturais e funcionais de hidrofobinas, poderia otimizar a prospecção, automatizando a detecção de novas classes e reduzindo a dependência de anotações manuais. Essa evolução metodológica permitiria explorar sistematicamente a diversidade de hidrofobinas em genomas fúngicos, acelerando sua aplicação em setores industriais.

5.2 Engenharia genética

A produção de proteínas recombinantes em fungos filamentosos, como *A. oryzae*, tem evoluído significativamente com o avanço de técnicas de biologia molecular, como CRISPR-Cas9, e a otimização de plataformas genéticas. Neste contexto, o presente trabalho alinha-se a esforços recentes para aprimorar a eficiência de expressão e secreção de proteínas anfipáticas, como as hidrofobinas, em *A. oryzae*. A seguir, discutem-se os resultados à luz da literatura, destacando avanços metodológicos e desafios persistentes.

5.3 Evolução das técnicas de transformação em *Aspergillus*

Estudos pioneiros, como o de Dersch, destacaram métodos clássicos de transformação em *Aspergillus*, como protoplastos e eletroporação, que eram limitados por baixas taxas de eficiência e integração aleatória no genoma. Essas técnicas, embora úteis, não permitiam edições precisas, o que comprometia a estabilidade das cepas (FLEIßNER; DERSCH, 2010).

A introdução do CRISPR-Cas9 revolucionou o campo, possibilitando edições genômicas direcionadas, como demonstrado por Li (YANGYANG LI et al.2024) em *A. niger*, que integrou múltiplos genes em loci específicos usando *arrays* de sgRNA-tRNA e DNAs doadores com braços homólogos longos (>1.000 pb). Essa abordagem aumentou a eficiência de integração em até 80%, superando métodos tradicionais baseados em NHEJ (junção não homóloga), que frequentemente resultam em inserções instáveis (LI et al., 2024). No presente trabalho, a estratégia de integração única no locus COSI de *A. oryzae* (via CRISPR-Cas9) mostrou-se eficaz para a expressão das proteínas, com secreção robusta, sugerindo que loci únicos bem

selecionados podem ser suficientes para proteínas menores e homólogas, como as anfipáticas.

5.4 Impacto do sistema de promotores na expressão

A escolha do promotor é crítica para a produção de proteínas recombinantes. Jarczynska utilizou o promotor *pgdA* na plataforma DIVERSIFY, obtendo alta expressão de proteínas heterólogas em *A. oryzae*. Contudo, no presente estudo, optou-se pelo promotor *pglaA* (associado à via de secreção de glucoamilase), que demonstrou desempenho comparável, com secreção eficiente, principalmente da RoIA. Essa diferença ressalta a flexibilidade de *A. oryzae* em adaptar-se a diferentes promotores, desde que acoplados a sistemas de sinalização adequados (JARCZYNSKA et al., 2021). Por outro lado, a baixa expressão da proteína TUB (20 kDa) pode estar relacionada às características intrínsecas da proteína, um desafio também observado por Nie (NIE et al.2025) na expressão de lipases em *A. niger*.

5.5 Comparação com produção de anticorpos em *A. oryzae*

O trabalho de Huynh demonstrou o potencial de *A. oryzae* para produção de anticorpos terapêuticos, como o adalimumab, utilizando CRISPR-Cas9 para deletar genes de hipermanosilação. Apesar da redução na ligação a FcγRIIIa, o estudo validou a viabilidade do fungo para proteínas complexas (FAN; LI, 2018). No presente estudo, a alta secreção de RoIA (uma proteína pequena de ~16 kDa) e os resultados promissores de TUB sugerem que *A. oryzae* pode ser adaptado para uma gama diversificada de biomoléculas, desde que otimizados elementos como promotores, peptídeos sinal e loci de integração. A ausência de contaminação significativa por proteínas endógenas no SDS-PAGE (Figura 8) reforça a eficácia do sistema de secreção utilizado, superando desafios reportados por Nie (NIE et al.2025), onde proteínas de fundo persistiam mesmo após deleções genéticas.

5.6 Limitações e perspectivas futuras

Embora a integração única no locus COSI tenha sido bem-sucedida para RoIA, a expressão subótima de TUB indica que proteínas algumas proteínas podem exigir estratégias de integração diferentes, como por exemplo a multiplex, como as

desenvolvidas por Li (FAN; LI, 2018). Além disso, a ausência de modificações pós-traducionais específicas em *A. oryzae* devem ser um ponto de atenção para aplicações terapêuticas, exigindo edições adicionais, como as realizadas por Huynh (HUYNH et al.2020). Futuros estudos poderiam explorar a combinação de múltiplos loci (e.g., *glaA* e *amyA*) e a deleção de proteases endógenas para maximizar a produção.

6 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo a prospecção, clonagem e expressão de proteínas anfipáticas recombinantes, especificamente hidrofobinas, utilizando ferramentas bioinformáticas integradas e técnicas avançadas de engenharia genética em *Aspergillus oryzae*. Através do desenvolvimento da ferramenta CPHunter, da análise de redes de similaridade de sequência (SSN) e da triagem de características estruturais e funcionais, foi possível identificar e selecionar candidatos proteicos com alto potencial biotecnológico.

A estratégia de clonagem baseada na substituição de cassete de expressão via CRISPR-Cas9 mostrou-se eficaz e adaptável ao sistema filamentoso, resultando em cepas recombinantes funcionais com expressão confirmada. Esses resultados representam uma contribuição relevante para o uso racional de bioinformática aplicada à expressão heteróloga de proteínas de interesse, além de validar *A. oryzae* como uma biofábrica promissora para proteínas com propriedades interfaciais.

As evidências obtidas neste estudo sustentam o uso de hidrofobinas recombinantes como ferramentas tecnológicas em diferentes frentes da biotecnologia, incluindo nanotecnologia, engenharia de materiais e farmacologia. A abordagem adotada reforça a importância da integração entre análise *in silico*, técnicas de edição genômica e bioprocessos na construção de soluções sustentáveis.

Como perspectiva futura, propõe-se o avanço da caracterização funcional e biofísica das proteínas produzidas, além da aplicação prática em sistemas de liberação de fármacos e biorremediação. Espera-se que este trabalho sirva de base para o desenvolvimento de novas estratégias de biotecnologia baseada em proteínas de superfície, fomentando a bioeconomia e a inovação científica nacional.

REFERÊNCIAS:

BALL, S. R.; KWAN, A. H.; SUNDE, M. Hydrophobin rodlets on the fungal cell wall. ***Current Topics in Microbiology and Immunology***, v. 425, p. 29–51, 2020. Disponível em: https://doi.org/10.1007/82_2019_186. Acesso em: 24 abr. 2025.

BERGER, B. W.; SALLADA, N. D. Hydrophobins: Multifunctional biosurfactants for interface engineering. ***Journal of Biological Engineering***, v. 13, n. 1, 2019. Disponível em: https://doi.org/10.1186/s13036-018-0136-1. Acesso em: 24 abr. 2025.

CHEN, Z., DUAN, R., XIAO, Y., WEI, Y., ZHANG, H., SUN, X., WANG, S., CHENG, Y., WANG, X., TONG, S., YAO, Y., ZHU, C., YANG, H., WANG, Y., & WANG, Z. (2022). Biodegradation of highly crystallized poly(ethylene terephthalate) through cell surface codisplay of bacterial PETase and hydrophobin. ***Nature Communications***, 13(1). https://doi.org/10.1038/s41467-022-34908-z

EL-GEBALI, S. et al. The Pfam protein families database in 2019. ***Nucleic Acids Research***, v. 47, n. D1, p. D427–D432, 2019. Disponível em: https://doi.org/10.1093/nar/gky995. Acesso em: 24 abr. 2025.

FAN, Gaowei; LI, Jinming. *Engineering antibodies for the treatment of infectious diseases*. In: FAN, Gaowei; LI, Jinming (ed.). **Recombinant antibodies for infectious diseases**. Cham: Springer, 2018. p. 207–220. (Advances in Experimental Medicine and Biology, v. 1053). Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-72093-0_13. Acesso em: 21 maio 2025.

FLEIßNER, A.; DERSCH, P. Expression and export: recombinant protein production systems for *Aspergillus*. ***Applied Microbiology and Biotechnology***, v. 87, p. 1255–1270, 2010. doi:10.1007/s00253-010-2672-6.

GLIER, A.; TRINDADE, R. S. O emprego de polímeros em revestimentos hidrofóbicos, super-hidrofóbicos e autolimpantes: uma revisão da literatura. ***Disciplinarum Scientia - Ciências Naturais e Tecnológicas*****,** v. 21, n. 2, p. 59–74, 2020. Disponível em: https://doi.org/10.37779/nt.v21i2.3481. Acesso em: 24 abr. 2025.

GUO, C. et al. Biodegradation of polyethylene terephthalate (PET) plastic by PETase: Structural features, catalytic mechanisms, and optimization strategies. ***Biotechnology Advances***, v. 56, 2022. Disponível em: https://doi.org/10.1016/j.biotechadv.2021.107888. Acesso em: 24 abr. 2025.

HATTAB, G.; WARSCHAWSKI, D. E.; MONCOQ, K.; et al. Challenges Associated With the Formation of Recombinant Protein in *Escherichia coli*. ***Frontiers in Bioengineering and Biotechnology*****,** v. 9, art. 630551, 2021. doi:10.3389/fbioe.2021.630551.

HORNUNG, B. V. H., & TERRAPON, N. (2023). An objective criterion to evaluate sequence-similarity networks helps in dividing the protein family sequence space***. PLoS Computational Biology,*** 19(8 August). https://doi.org/10.1371/journal.pcbi.1010881

JIN, F.-J.; HU, S.; WANG, B.-T.; JIN, L. Advances in Genetic Engineering Technology and Its Application in the Industrial Fungus *Aspergillus oryzae*. ***Frontiers in Microbiology***, v. 12, art. 644404, 23 fev. 2021. doi:10.3389/fmicb.2021.644404.

JENSEN, L. J., JULIEN, P., KUHN, M., VON MERING, C., MULLER, J., DOERKS, T., & BORK, P. (2008). EGGNOG: Automated construction and annotation of orthologous groups of genes. ***Nucleic Acids Research***, 36(SUPPL. 1). https://doi.org/10.1093/nar/gkm796

KULKARNI, S. S., NENE, S. N., & JOSHI, K. S. (2022). Identification and characterization of a hydrophobin Vmh3 from Pleurotus ostreatus. ***Protein Expression and Purification***, 195–196, 106095. https://doi.org/10.1016/J.PEP.2022.106095

LAM, Wai-Hei et al. Talaromyces marneffei Mp1 Protein, a Novel Virulence Factor, Carries Two Arachidonic Acid-Binding Domains To Suppress Inflammatory Responses in Hosts**. *Infection and Immunity***, [s.l.], v. 87, n. 4, p. e00679-18, abr. 2019. Disponível em: https://journals.asm.org/journal/iai. Acesso em: 17 maio 2025.

LITTLEJOHN, Karen A.; HOOLEY, Paul; COX, Phil W. Bioinformatics predicts diverse Aspergillus hydrophobins with novel properties. *Food Hydrocolloids*, [S. l.], v. 27, n. 2, p. 503–516, jun. 2012. Elsevier. Disponível em: https://doi.org/10.1016/j.foodhyd.2011.08.018. Acesso em: 21 maio 2025.

LOPEZ, C. (2024). ***Protocol for Gibson Assembly v1.*** https://doi.org/10.17504/protocols.io.n92ld8m48v5b/v1

MIYAZAWA, K.; YOSHIMI, A.; SANO, M.; TABATA, F.; SUGAHARA, A.; KASAHARA, S.; et al. Improved recombinant protein production in *Aspergillus oryzae* lacking both α-1,3-glucan and galactosaminogalactan in batch culture with a lab-scale bioreactor. ***Journal of Bioscience and Bioengineering*****,** v. 133, p. 39–45, set. 2021. doi:10.1016/j.jbiosc.2021.09.010.

NIE, Hongmei et al. Heterologous expression of *Candida antarctica* lipase B in *Aspergillus niger* using CRISPR/Cas9-mediated multi-gene editing. *Biotechnology and Bioengineering*, [S. l.], v. 2025, p. 1–10, 2025. Wiley Periodicals LLC. Disponível em: https://doi.org/10.1002/bit.29000. Acesso em: 21 maio 2025.

OHTAKI, S., MAEDA, H., TAKAHASHI, T., YAMAGATA, Y., HASEGAWA, F., GOMI, K., NAKAJIMA, T., & ABE, K. (2006). Novel hydrophobic surface binding protein, HsbA, produced by Aspergillus oryzae. ***Applied and Environmental Microbiology,*** 72(4), 2407–2413. https://doi.org/10.1128/AEM.72.4.2407- 2413.2006

OHTAKI, S. et al. Genome sequencing and analysis of Aspergillus oryzae. ***Nature***, v. 438, p. 1157–1161, 2005. Disponível em: https://doi.org/10.1038/nature04300. Acesso em: 24 abr. 2025.

PUNT, P. J.; VAN DEN BIJEZEN, N.; CONESA, A.; et al. Filamentous fungi as cell factories for heterologous protein production. ***Trends in Biotechnology***, v. 20, n. 5, p. 200–206, maio 2002. doi:10.1016/S0167-7799(02)01933-9.

PUSPITASARI, N., Tsai, S. L., & Lee, C. K. (2021). Fungal Hydrophobin RolA Enhanced PETase Hydrolysis of Polyethylene Terephthalate. ***Applied Biochemistry and Biotechnology***, 193(5), 1284–1295. https://doi.org/10.1007/s12010-020-03358-y

OBERG, N., ZALLOT, R., & GERLT, J. A. (2023). EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. ***Journal of Molecular Biology***, 435(14). https://doi.org/10.1016/j.jmb.2023.168018

SCIENCEDIRECT TOPICS. *Aspergillus oryzae* – an overview. In: ***Pharmacology, Toxicology and Pharmaceutical Science*****.** Disponível em: https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/aspergillus-oryzae. Acesso em: 19 abr. 2025.

TANAKA, T., TERAUCHI, Y., YOSHIMI, A., & ABE, K. (2022). Aspergillus Hydrophobins: Physicochemical Properties, Biochemical Properties, and Functions in Solid Polymer Degradation***. In Microorganisms*** (Vol. 10, Issue 8). MDPI. https://doi.org/10.3390/microorganisms10081498

TURGUT, B. A., & ÖRTÜCÜ, S. (2023A). Hydrophobins: The Amphiphilic Proteins Produced in Filamentous Fungi. ***In Original Article Eurasian Mol Biochem Sci*** (Vol. 2, Issue 1). https://dergipark.org.tr/tr/pub/ejmbs

U.S. FOOD AND DRUG ADMINISTRATION (FDA). GRAS Notice No. GRN 000829 – Powdered *Aspergillus oryzae* grown with one or more added minerals. ***FDA***, 9 dez. 2019. Disponível em: https://www.fda.gov/media/132948/download. Acesso em: 19 abr. 2025.

WANG, L. et al. Application of hydrophobins in nanomedicine: A review. ***Advanced Drug Delivery Reviews***, v. 182, p. 114114, 2022. Disponível em: https://doi.org/10.1016/j.addr.2021.114114. Acesso em: 24 abr. 2025.

WANG, Y. et al. A novel hydrophobin fusion protein for enhanced drug delivery in cancer therapy. ***Biomaterials Science***, v. 10, n. 2, p. 402–411, 2022. Disponível em: https://doi.org/10.1039/D1BM01234A. Acesso em: 24 abr. 2025.

XU, H.; WANG, S.; LI, H.; et al. Aspergillus: ***A Powerful Protein Production Platform. Catalysts***, v. 10, n. 9, art. 1064, 8 set. 2020. doi:10.3390/catal10091064.

YANG, L. et al. Fungal hydrophobins: A versatile protein family for surface modification and drug delivery. ***Biotechnology Journal***, v. 17, n. 5, p. e2100556, 2022. Disponível em: https://doi.org/10.1002/biot.202100556. Acesso em: 24 abr. 2025.

ANEXO A

: Gla Signal Peptide

Boldaaai: Primers

: Natural Signal peptide

: Stop

: HisTag

Introjjiiin: Introns

RolA: 861 COMPLETE – RNAm: 557 - chromosome 6 (Puc57+cass+rola: 5901 pb**)**

>AB094496.1 Gene RolA de Aspergillus oryzae para hidrofobina putativo, cds completos

TGGCAAATGTGATTTCCAAGATGCAGTTCTCCGTCGCCGCTGTTCTTGCTCTGG
CTACTGCCGTTGCCGCT**CTTCCTCCTGCCTCTGGC**ACTGGCGCTGGCCAGCAA
GTCGGACACTCCAAGAACGACTTCCCTCTCCCTAAGGAGTTGACCACCAAGCAG
GCCGCCGACAAGTGTGGTGACCAGGCTCAGCTCACCTGCTGCAACAAGACCGT
CAAGACCGGTGACTTCACCCAGGTTGAGGAGGGTCTCCTTGCTGGCCTCCTCT
CCAACCTCCTCGGTGCCGGACAGGGCTCCCAGGGTCTTGGTCTCTTGGATGAG
TGCACCAACATCCCTGTTATCCCCATCAGTAAGTATCTAGCAACATGAGATATAG
CCAATGGCATACTAACAAAGCTTAGTCTCCATCGCCTCTCCTCAGGAGAAGTGC
AAGCAGCCCATCTCTTGCTGCCAGAACACCAAGTCCAGCGCCGTAAGTAGAATC
AATGCACAATATACCCAGCTTTCACTAACAATCACAGGATGGCGACCTCGTCGG
TATTGGTCTTCCTTGCATC**GCTCTCGGCTCTCTCCTG**TAACACCATCATCACCAC
CACTA

PRIMER F: TGGCAAATGTGATTTCCAAGCTTCCTCCTGCCTCTGGC

PRIMER R: TAGTGGTGGTGATGATGGTGCAGGAGAGAGCCGAGAGC

H3 complete 1616 – RNAm (CLONADO 6957 PB)

>NC_036438.1:911897-914025 Aspergillus oryzae RIB40 DNA, chromosome 4

TGGCAAATGTGATTTCCAAG**ATGAAGGTTTCTACCTTGTTGACCACCCTTTGCG**
TACTGGGTGTTGCTAGTTCGGCGGATCCAACCGCACAATGCGATAACGGACCC
GTTCAATGCTGCGCAACGGTCGGTTTACCGACTGACAGTGTAGTCTCACCTTTA
CTTGGCTTGCTCGGAGTAGTTGTTCCAGACATGAGCACTCCAGTTGGATTGACC
TGTATGTCTCGATATTCTCCTTTGCACCTTCGAAAACTAATAAGTCACCTTTTCAA

GGCAATCCTATTGTCCAGGGTGGATCTTGCCCAGGCCACCCCGTTTGCTGTACA
GGAAATGTAGGCCATGCTTCTGCGACACATTGTTCCAGCTACCTAATGTCCATC
GGAAATGATTACTGACACGAGTACAGAATTTCAATGGTCTCCTAGCCCTTGGCT
GTACACCCGTCACGTTATCTACATGAAAACAGTAAAAAGAAAGCAAAGGCAGGC
ACGTTTCAGTTAGTATTTCGATCTGGTTGGGGGAGAGGGAGGCCAAAATGTCTC
GGACGAAAATTGATTACAAGTGTACGTTGAGCAGAAATGTTCAAGGTGGTGATG
ACCGTATAGAAACATGTCTTATGATGCGATGTGGCCTCGCGAGCGGTTGTTGAT
TTGTTATTACTTCAATCTCAGGCAACCAGGTAGCAGATGTCCAGCTAATGGCATT
AGACGATTATCTACAAGTGGTCTTAGTGTCTTGTATCTTGTATTACAGGAGACTA
CAACTAGGAAATCCTTGGTCATGCCTAGAGTTATTAACGGGTTACGGCTCAGGT
TCAAGATATAAGACCCGGCCCGAGCCCGCGCAGATTTAGGTTCCCCGCGCTGG
AGCTAAATAGGTATAGCATTACTTAGGCCTAGTGTACACCCCTAGACTAATTTTG
CAAGCTCTAATCTGTATATAAATGTAAAAGCAATTCTAGAAAATCTATGGTTTTAA
CCAGGTGATAACAACAAAAAGCCGATGGCTTCCTCCGCAATCGGGTCGGGTTTT
GCGGGGAACCCGAACAACTCTAGCTACAAGTATTAAGCACCCCGTTCCTCTTAC
TTCTTTCGTCTCAGAAAACACTCGCTCATGGTCCGAAACAGAAGTCCAATTAGGA
CCCCCGGCTGATATGAAGCGCCGGAGCAATGTCAATCTTCTCTCAAGCTTTTAA
ACAATTGATGCCAAAACCCTGAAATATCTTCTGCAACAACATAATCACACTCATTT
ATCCGTTCTCTCAACTCGTCAAGAAGATCCCTACGTGCCTGCAGACCCCACGAA
GATGTCCTCCAGGGAGCTGCTCTCAAATCCCGCACCAAAGCTATCACCGCCCC
AAACAACATTGCAAAGCTATCCAAGGACCATGCAACTGTATTACGCCAAGGGCA
GTCATAGATATAGAAACCGGCTCCACCGTACCGGACATCATCTGGTTCCCAAAG
CCATACATCACCGTATTCTGCGTCTACCATTACCATACAGTCGTTGACATTTCCG
ATGTGGAAATAGATGTCACTGCACTGGGCTCGCGAACAAGGTATCTCATGTATT
TCAAGTTCATTGTCGTCGGTAAGGATGTAGTGTTTCGGACT**ACTCTTGGC**
TTCTTGGCTTACACCATCATCACCACCACTA

PRIMER F: TGGCAAATGTGATTTCCAAGGCGGATCCAACCGCACAATG
PRIMER R: TAGTGGTGGTGATGATGGTGAGCCAAGAAGCCAAGAGTAGTC

TUB

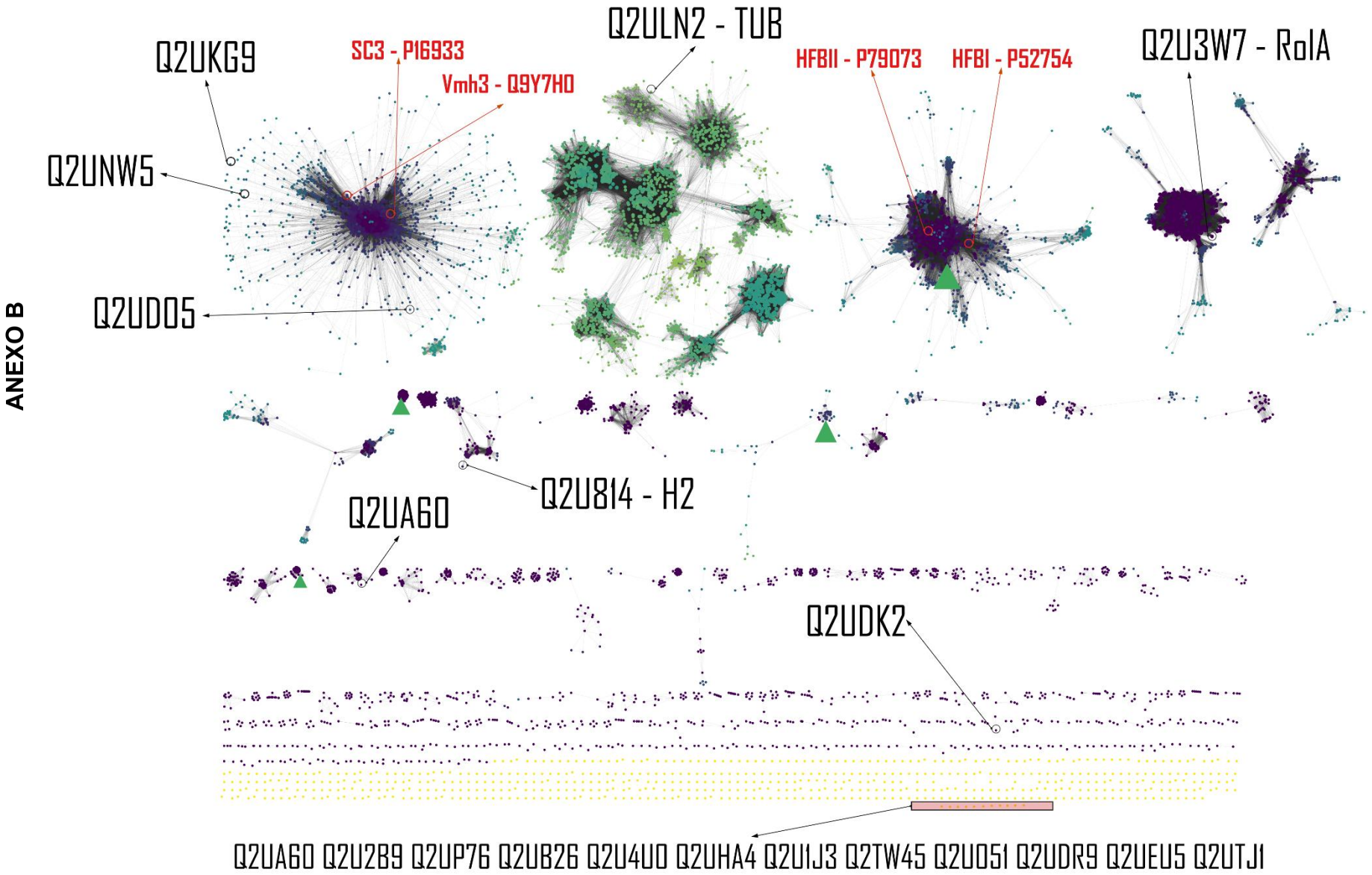
>2 dna:chromosome chromosome 2 : ASM18445v3:2:3019867:3020696:1 RNAm (830) COMPLETE (reverso complementar da referência do banco de dados)

(clonada cass 6225 bp)

TGGCAAATGTGATTTCCAAG**ATGCAACGCTCGCGAATTAA**CTTTCGTTAGCTTG
TCTAAAATCTTCCACCTGAGGCAGGCTGCATGAATCCAACAAATCAGGAATCTT
GTGTGCTTCTTTCAAAGAAGCTGCATGGCAGTCGTGGGTCGATCTGGCCAGAAA
AAGCCTCAAGAATGCGATACGGCGAAAATCGAGAAATCGAGTCATTCGGTGGAA
GGAGCTAAGGGATCCGCTTTACACTCACTCAGGGCTCACGAGATTATGATCCTT
TAGACGATATAGCTGCTCTTCACTAGGCACAGGAGTGCTCGGTTTCAATTATCTT
CGCCGGAGGGCATTGTCCGACTCAAGTGCACGGGCCGCTACACCTATTCTTCG
CCGCGATGCAGCTAAAGTTCAGGGCGATATAACGCAAAAGATCGAACCGCAAAT
TAACACCCTTTACAATGATGTCCGTGGTTTCCCTACTAGTGGTTTAACCGGAGCA
ATGACTATTCGCAGCGACTTGCAGAGCCTTGCTACCACCGTGAATGATGCGACG
GCCGATGTCAAATCCACCGGCTCCTTGGACACTCCATCTGGAACTACTATTCTT
GCCGACATCCAATCGCTGATGCCGACGTCCTTGGTTACATTAACTCACGTTGGA
GCCGAAGCACCTGCATGGGAAGATATTCAAGGGGGGCCAGCTCTGATTCTTAG
CGATCTCCGGTCTCTCAAAACGGCCTTGGACAATTTCGCCAACGCTCTGATATC
GAACGAGCCCCTTCTTCTACAAGCCAAGGCTCTCGCGATCAAGACTCAGATTGA
CGGTGGTCTCGACATTG**CAATCGCCCCCTACTCCGTA**TAGCACCATCATCACCA
CCACTA

PRIMER F: TGGCAAATGTGATTTCCAAGATGCAACGCTCGCGAATTAAC

PRIMER R: TAGTGGTGGTGATGATGGTGTACGGAGTAGGGGGCGATTG

SSN - HIDROFOBINAS - TODAS CLASSES

ANEXO B