



**Universidade de Brasília**  
Instituto de Ciências Biológicas  
Departamento de Biologia Celular  
Programa de Pós-Graduação em Biologia Molecular

# **Estratégias para Integração Múltipla de Cassetes de Expressão no Genoma de *Komagataella phaffii***

Maritza Ocampo Betancur

**Orientador:** Prof. Fernando Araripe Gonçalves Torres

**Coorientadora:** Dr<sup>a</sup> Viviane Castelo Branco Reis

Brasília, julho de 2017

Universidade de Brasília  
Instituto de Ciências Biológicas  
Departamento de Biologia Celular  
Programa de Pós-Graduação em Biologia Molecular

# Estratégias para Integração Múltipla de Cassetes de Expressão no Genoma de *Komagataella phaffii*

**Maritza Ocampo Betancur**

Orientador: Prof. Dr. Fernando Araripe Gonçalves Torres

Coorientadora: Dr<sup>a</sup> Viviane Castelo Branco Reis

*Tese de doutorado apresentada ao departamento de Biologia Celular do Instituto de Ciências Biológicas da Universidade de Brasília como requisito parcial para a obtenção do título de Doutora em Biologia Molecular.*

Brasília, 2017

Trabalho realizado no Laboratório de Biologia Molecular, Departamento de Biologia celular,  
Instituto de Ciências Biológicas, Universidade de Brasília, sob a orientação do professor  
Fernando Araripe Gonçalves Torres com apoio financeiro da CAPES.

**Banca Examinadora**

Professora Elis Cristina Araújo Eleuthério  
Universidade Federal do Rio de Janeiro

Professora Élide Geralda Campos  
Universidade de Brasília

Professora Lídia Maria Pepe de Moraes  
Universidade de Brasília

*“Sábio é o ser humano que tem coragem de ir  
diante do espelho da sua alma para reconhecer  
seus erros e fracassos e utilizá-los para plantar  
as mais belas sementes no terreno de sua  
inteligência”*

Augusto Cury

*Dedico esta conquista a todas aquelas pessoas que acreditaram em mim e que me deram forças para alcançar minha meta. Em especial aos meus pais por toda a entrega e dedicação para minha formação.*

## AGRADECIMENTOS

Em primeiro lugar agradeço à vida pela oportunidade que me deu de continuar minha formação em outro país no qual pude aprender tantas coisas novas.

A toda minha família pelo apoio para empreender meu sonho de voar longe. Aos meus pais por tanto amor, por suas lições de vida, pelos conselhos e paciência nos momentos em que quis desistir. À minha irmã pela amizade e por alegrar meus dias mesmo estando longe. A todos meus tios e primos pelo apoio emocional.

Ao professor Fernando Araripe pela oportunidade e pela confiança depositada em mim. Além de um excelente orientador, foi como um pai sempre preocupado com meu bem-estar.

À Viviane Reis pela orientação durante todo o projeto, pelos ensinamentos, pela disponibilidade para tirar dúvidas e por ficar do meu lado nas discussões científicas com o Fernando.

À professora Lídia Pepe Moraes pelo carinho, pelos ensinamentos e pela ajuda no planejamento experimental.

Ao professor André Nicola pela ajuda com os ensaios de citometria de fluxo.

A todos meus amigos na Colômbia pelos momentos compartilhados nas férias e pela companhia através da internet. Em especial às amigas do colégio Elizabeth, Christina e Karla porque apesar da distância nunca houve uma amizade mais forte. Ao David por estar sempre disposto a me escutar, aconselhar e tirar dúvidas.

Aos amigos Daniel e Lucas pela convivência. Vocês foram minha família no Brasil e sempre estarão no meu coração.

A todos os colegas do laboratório Biomol, em especial aos amigos do Lab 2: Dani, Carol, Yasmin, Diego, Roberta, Ieda, Chris, Myrna, Francisco, Vanessa, Tiago, Marciano e a professora Janice; assim como aos que já não estão mais no lab: Daniel, Danuza, Neumara,

Lucas, Túlio e Juliana. Agradeço especialmente a minha companheira de bancada e grande amiga Luiza, pelas respostas a minhas perguntas filosóficas e pela ajuda na hora de tomar decisões. Viva a engenharia!!

À Ana Laura, que mais que colega de laboratório, foi como uma irmã. Obrigada pela companhia, pelos favores, pelas festas, por tudo!!

A todas as pessoas que passaram por minha vida ao longo do tempo que durou meu doutorado. Cada uma delas foi importante demais no seu respectivo momento. Principalmente agradeço a minha querida amiga argentina Carolina por todo o carinho e pela convivência durante o tempo que moramos juntas. Ao Daniel pela companhia, por tantas horas de risos no Skype e por todos os momentos de alegria. E aos amigos colombianos que conheci aqui no Brasil que me fizeram sentir um pouco mais perto do meu país.

A todos os professores que participaram da minha formação.

Aos membros da banca examinadora, pela disponibilidade e paciência para a correção do meu trabalho.

À Capes pelo apoio financeiro.

A todos que de alguma forma colaboraram para a realização deste trabalho.

*“As grandes ideias surgem da observação dos pequenos detalhes”*

Augusto Cury

## INDICE

Lista de figuras	i
Lista de tabelas	v
Resumo	vi
Abstract	vii
<b>1 Introdução</b>	<b>1</b>
1.1 <i>Komagataella phaffii</i> e expressão heteróloga	1
1.2 Integração múltipla	9
1.3 Marcas auxotróficas defectivas	14
<b>2 Objetivos</b>	<b>16</b>
2.1 Objetivo geral	16
2.2 Metas	16
<b>3 Estratégia</b>	<b>17</b>
<b>4 Material e Métodos</b>	<b>18</b>
4.1 Material	18
4.1.1 Células	18
4.1.2 Meios de cultura	18
4.1.3 Soluções	20
4.1.4 Kits	27
4.1.5 Oligonucleotídeos ( <i>primers</i> )	28
4.1.6 Marcadores de massa molecular para DNA	29
4.1.7 Enzimas de restrição	29
4.1.8 Vetores	30
4.2 Métodos	32
4.2.1 Análise bioinformática	32



4.2.2	Amplificação por PCR	32
4.2.3	Preparação de células bacterianas competentes	33
4.2.4	Manipulação do DNA	34
4.2.5	Construção de plasmídeos	34
4.2.6	Preparação de DNA plasmidial	35
4.2.7	Transformação de leveduras por eletroporação	36
4.2.8	Extração de DNA total de leveduras	36
4.2.9	Southern blot	37
4.2.10	Eletroforese em campo pulsado	37
4.2.11	Análise de fluorescência no <i>Typhoon scanner</i>	38
4.2.12	Citometria de fluxo	38
4.2.13	Eletroforese em gel de poliacrilamida	39
4.2.14	Análise de fluorescência em gel de poliacrilamida não desnaturante	39
4.2.15	Detecção da atividade amilase em placa	40
4.2.16	Detecção da atividade amilase em meio líquido	40
4.2.17	Determinação de biomassa	40
4.2.18	Ensaio enzimático de degradação de amido	41
4.2.19	Detecção de atividade amilolítica em gel de poliacrilamida – zimograma	41
4.2.20	Teste de estabilidade genética	42
4.2.21	Avaliação do crescimento em meio líquido	43
4.2.22	Análise de dados	43
<b>5</b>	<b>Resultados e Discussão</b>	<b>44</b>
5.1	Identificação de sequências repetidas no genoma de <i>K. phaffii</i>	44
5.2	Integração múltipla do gene repórter EGFP	51
5.2.1	Construção do vetor pKld-GFP	51

5.2.2 Construção dos vetores pBSChr2ld-GFP, pBSChr3ld-GFP, pBS5Sld-GFP e pPCVNTSld-GFP	54
5.2.3 Transformação de <i>K. phaffii</i> M12	59
5.2.4 Determinação do número de cópias	66
5.2.5 Produção de GFP	80
5.2.6 Estabilidade genética	86
5.3 Integração múltipla do gene <i>AMY1</i>	90
5.3.1 Construção dos vetores pBSChr2ld-AMY, pBSChr3ld-AMY, pBS5Sld-AMY e pPCVNTSld-AMY	91
5.3.2 Transformação de <i>K. phaffii</i> M12	95
5.3.3 Determinação do número de cópias	99
5.3.4 Produção de amilase	103
5.3.5 Estabilidade genética	112
5.4 Correlação entre o número de cópias e a produção de proteínas	113
<b>6 Conclusões e Perspectivas</b>	118
<b>7 Referências</b>	119
<b>8 Anexos</b>	127
Anexo 1: Sequência sonda PGK1	127
Anexo 2: Sequência sonda EGFP	127
Anexo 3: Padrão dos marcadores de massa molecular usados	128
Anexo 4: Mapa físico do vetor pKld-GFP	129
Anexo 5: Mapa físico do vetor pBSChr2ld-GFP	130
Anexo 6: Mapa físico do vetor pBSChr3ld-GFP	131
Anexo 7: Mapa físico do vetor pBS5Sld-GFP	132
Anexo 8: Mapa físico do vetor pPCVNTSld-GFP	133
Anexo 9: Mapa físico do vetor pBSChr2ld-AMY	134
Anexo 10: Mapa físico do vetor pBSChr3ld-AMY	135

Anexo 11: Mapa físico do vetor pBS5SId-AMY	136
Anexo 12: Mapa físico do vetor pPCVNTSId-AMY	137
Anexo 13: Mapa físico do vetor pkL2-AMY	138
<b>9 Patente e artigos publicados</b>	<b>139</b>

## LISTA DE FIGURAS

<b>Figura 1.</b> Número de publicações usando a levedura <i>K. phaffii</i> desde 1980 até o presente ano	3
<b>Figura 2.</b> Esquema de um vetor de expressão para <i>K. phaffii</i>	6
<b>Figura 3.</b> Estratégias para gerar clones multicópia em <i>K. phaffii</i>	11
<b>Figura 4.</b> Representação esquemática do <i>locus</i> do rDNA em <i>K. phaffii</i>	13
<b>Figura 5.</b> Cálculo da velocidade máxima específica de crescimento ( $\mu_{\text{máx}}$ )	43
<b>Figura 6.</b> Sequência alvo de integração no cromossomo 2 de <i>K. phaffii</i>	47
<b>Figura 7.</b> Sequência alvo de integração no cromossomo 3 de <i>K. phaffii</i>	48
<b>Figura 8.</b> Sequência alvo de integração na região NTS do rDNA de <i>K. phaffii</i>	49
<b>Figura 9.</b> Localização da sequência do rDNA 5S nos cromossomos de <i>K. phaffii</i>	50
<b>Figura 10.</b> Estratégia para a construção do vetor pKld-GFP	52
<b>Figura 11.</b> Clonagem do alelo <i>leu2-d</i> no vetor pPICK2	53
<b>Figura 12.</b> Clonagem do gene EGFP no vetor pKld	54
<b>Figura 13.</b> Sequências repetidas sintetizadas	55
<b>Figura 14.</b> Estratégia para a construção de cassete de expressão com sequências repetitivas	56
<b>Figura 15.</b> Sítios de anelamento dos <i>primers</i> para amplificar o cassete PGK-EGFP-leud	57
<b>Figura 16.</b> Confirmação da clonagem do fragmento PGK-EGFP-leu2d nos plasmídeos com sequências repetidas	58
<b>Figura 17.</b> Resultado da transformação de <i>K. phaffii</i> M12 com o vetor pKld-GFP	59
<b>Figura 18.</b> Cinética de crescimento dos clones transformados com o vetor pKld-GFP	60
<b>Figura 19.</b> Expressão intracelular de EGFP em clones transformados com o vetor pKld-GFP	61

<b>Figura 20.</b> Expressão intracelular de EGFP em clones transformados com os cassetes contendo sequências repetidas	62
<b>Figura 21.</b> Cinética de crescimento dos clones transformados com o cassete Chr2ld-GFP	63
<b>Figura 22.</b> Cinética de crescimento dos clones transformados com o cassete Chr3ld-GFP	63
<b>Figura 23.</b> Cinética de crescimento dos clones transformados com o cassete 5Sld-GFP	64
<b>Figura 24.</b> Cinética de crescimento dos clones transformados com o cassete NTSld-GFP	64
<b>Figura 25.</b> Determinação do número de cópias dos clones transformados com o vetor pKld-GFP	67
<b>Figura 26.</b> Correlação entre a velocidade específica de crescimento e o número de cópias dos clones transformados com o vetor pKld-GFP	68
<b>Figura 27.</b> Determinação do número de cópias dos clones transformados com o cassete Chr2ld-GFP	69
<b>Figura 28.</b> Determinação do número de cópias do clone 1 transformado com o cassete Chr2ld-GFP	70
<b>Figura 29.</b> Determinação do número de cópias dos clones transformados com o cassete Chr3ld-GFP	71
<b>Figura 30.</b> Determinação do número de cópias dos clones 3 e 4 transformados com o cassete Chr3ld-GFP	72
<b>Figura 31.</b> Determinação do número de cópias dos clones transformados com o cassete 5Sld-GFP	73
<b>Figura 32.</b> Determinação do número de cópias dos clones transformados com o cassete NTSld-GFP	74
<b>Figura 33.</b> Correlação entre a velocidade específica máxima de crescimento e o número de cópias dos clones transformados com os cassetes contendo sequências repetidas e o gene EGFP	75
<b>Figura 34.</b> Determinação do local de integração do cassete Chr2ld-GFP	77
<b>Figura 35.</b> Determinação do local de integração do cassete Chr3ld-GFP	78
<b>Figura 36.</b> Determinação do local de integração do cassete NTSld-GFP	79

<b>Figura 37.</b> Produção intracelular de GFP em clones transformados com pKld-GFP	81
<b>Figura 38.</b> Produção intracelular de GFP em clones transformados com o vetor pKld-GFP	82
<b>Figura 39.</b> Produção intracelular de GFP em clones transformados com os cassetes Chr2ld-GFP e Chr3ld-GFP	83
<b>Figura 40.</b> Produção intracelular de GFP em clones transformados com os cassetes 5Sld-GFP e NTSld-GFP	83
<b>Figura 41.</b> Estabilidade genética dos clones transformados com o vetor pKld-GFP	86
<b>Figura 42.</b> Estabilidade genética dos clones transformados com os cassetes Chr2ld-GFP e Chr3ld-GFP	87
<b>Figura 43.</b> Estabilidade genética dos clones transformados com os cassetes NTSld-GFP e 5Sld-GFP	88
<b>Figura 44.</b> Estratégia para a construção dos vetores pBSChr2ld-AMY, pBSChr3ld-AMY, pBS5Sld-AMY e pPCVNTSld-AMY	91
<b>Figura 45.</b> Amplificação do gene <i>AMY1</i>	92
<b>Figura 46.</b> Clonagem do gene <i>AMY1</i> nos vetores com sequências repetidas	93
<b>Figura 47.</b> Clonagem do gene <i>AMY1</i> no vetor com sequências 5S	94
<b>Figura 48.</b> Clonagem do gene <i>AMY1</i> no vetor com marca <i>LEU2</i>	95
<b>Figura 49.</b> Formação de halos de hidrólise de amido em clones transformados com os cassetes contendo sequências repetidas e com o vetor pKL2-AMY	96
<b>Figura 50.</b> Cinética de crescimento dos clones transformados com o cassete Chr2ld-AMY	97
<b>Figura 51.</b> Cinética de crescimento dos clones transformados com o cassete Chr3ld-AMY	97
<b>Figura 52.</b> Cinética de crescimento dos clones transformados com o cassete 5Sld-AMY	98
<b>Figura 53.</b> Determinação do número de cópias dos clones transformados com o vetor pKL2-AMY	99
<b>Figura 54.</b> Determinação do número de cópias dos clones transformados com os cassetes Chr2ld-AMY e Chr3ld-AMY	101

<b>Figura 55.</b> Determinação do número de cópias dos clones transformados com o cassete 5Sld-AMY	102
<b>Figura 56.</b> Correlação entre a velocidade específica máxima de crescimento e o número de cópias dos clones transformados com os cassetes contendo sequências repetidas e o gene <i>AMY1</i>	103
<b>Figura 57.</b> Correlação entre o peso seco e a OD <sub>600</sub> da levedura <i>K. phaffii</i> M12	104
<b>Figura 58.</b> Cinética de produção de amilase em meio complexo na linhagem <i>K. phaffii</i> M12 e no clone contendo uma cópia do gene <i>AMY1</i>	105
<b>Figura 59.</b> Cinética de produção de amilase em meio complexo nos clones transformados com os cassetes contendo sequências repetidas	106
<b>Figura 60.</b> Cinética de produção de amilase em meio mínimo na linhagem <i>K. phaffii</i> X-33 e no clone contendo uma cópia do gene <i>AMY1</i>	107
<b>Figura 61.</b> Cinética de produção de amilase em meio mínimo nos clones transformados com os cassetes contendo sequências repetidas	108
<b>Figura 62.</b> Zimograma de clones selecionados produzindo amilase	111
<b>Figura 63.</b> Estabilidade genética dos clones transformados com os cassetes Chr2ld-AMY, Chr3ld-AMY e 5Sld-AMY	112
<b>Figura 64.</b> Correlação entre a fluorescência e o número de cópias para os clones transformados com o vetor pKld-GFP	113
<b>Figura 65.</b> Correlação entre a velocidade específica máxima de crescimento e a fluorescência para os clones transformados com o vetor pKld-GFP	114
<b>Figura 66.</b> Correlação entre a fluorescência e o número de cópias para os clones transformados com os cassetes contendo as sequências repetidas e o gene EGFP	114
<b>Figura 67.</b> Correlação entre a velocidade específica máxima de crescimento e a produção de GFP para os clones transformados com os cassetes com sequências repetidas	115
<b>Figura 68.</b> Correlação entre a atividade amilolítica e o número de cópias para os clones transformados com os cassetes contendo as sequências repetidas e o gene <i>AMY1</i>	115
<b>Figura 69.</b> Correlação entre a velocidade específica máxima de crescimento e a produção de amilase para os clones transformados com os cassetes com sequências repetidas	116

## LISTA DE TABELAS

<b>Tabela 1.</b> Alguns dos produtos no mercado ou na etapa final de desenvolvimento produzidos em <i>K. phaffii</i>	4
<b>Tabela 2.</b> <i>Primers</i> utilizados neste trabalho	28
<b>Tabela 3.</b> Enzimas de restrição usadas	29
<b>Tabela 4.</b> Vetores utilizados	30
<b>Tabela 5.</b> Fragmentos amplificados por PCR e condições da amplificação	32
<b>Tabela 6.</b> Sequências repetidas <i>in tandem</i> identificadas no genoma de <i>K. phaffii</i>	44
<b>Tabela 7.</b> Velocidade específica máxima de crescimento dos clones transformados com o vetor pKld-GFP	60
<b>Tabela 8.</b> Velocidade específica máxima de crescimento dos clones transformados com os cassetes Chr2ld-GFP, Chr3ld-GFP, 5Sld-GFP e NTSld-GFP	65
<b>Tabela 9.</b> Relação de número de cópias, produção relativa de GFP e velocidade de crescimento	84
<b>Tabela 10.</b> Velocidade específica máxima de crescimento dos clones transformados com os cassetes Chr2ld-AMY, Chr3ld-AMY e 5Sld-AMY	98
<b>Tabela 11.</b> Atividade amilolítica dos clones transformados com os cassetes contendo sequências repetidas	109
<b>Tabela 12.</b> Relação de número de cópias, produção relativa de amilase e velocidade de crescimento	110



## RESUMO

O aumento do número de cópias do gene heterólogo é uma das abordagens empregadas no melhoramento genético de *Komagataella phaffii* como sistema de expressão. O uso de marcas auxotróficas defectivas é uma estratégia para a obtenção de integrantes multicópia. Para garantir estabilidade mitótica, os cassetes de expressão são integrados no genoma por recombinação homóloga, portanto é necessário haver no vetor sequências genômicas que propiciem esses eventos de integração. O foco deste trabalho foi identificar sítios repetitivos no genoma de *K. phaffii* nunca antes testados como alvos de integração, em combinação com o uso da marca defectiva *leu2-d*, para aumentar a expressão heteróloga. Como alvos para integração múltipla foram avaliados os seguintes *loci*: rDNA 5S, região NTS do rDNA e regiões repetidas dos cromossomos 2 e 3. Um vetor contendo a marca *leu2-d* e o gene repórter EGFP (*Enhanced Green Fluorescent Protein*) foi construído e foi usado como plataforma para clonar as diversas sequências repetidas. Todas as construções foram utilizadas para transformar *K. phaffii* M12 (*leu2*). A determinação do número de cópias integradas do cassete de expressão confirmou a obtenção de clones multicópia com todas as construções. Até 78 cópias do cassete contendo a sequência repetitiva do cromossomo 2 foram integradas. Todos os clones multicópia apresentaram uma maior produção intracelular de GFP em comparação a um clone contendo uma única cópia do gene, chegando a se obter um aumento de 197 vezes com o clone contendo 78 cópias integradas do cassete.

Para validar a estratégia apresentada neste estudo, foi testada a produção extracelular da proteína repórter  $\alpha$ -amilase. Clones multicópia foram obtidos com as diferentes construções e até 43 cópias do cassete contendo a sequência repetitiva do cromossomo 2 foram integradas. A atividade amilolítica no sobrenadante confirmou a produção de amilase, que foi maior para os clones multicópia em comparação a um clone contendo uma cópia do gene, chegando a se obter um aumento de 4,6 vezes com o clone contendo 43 cópias integradas do cassete. A produção das duas proteínas avaliadas, demonstrou a possibilidade de obter clones multicópia que produzem uma maior quantidade da proteína recombinante. Contudo, foi observada perda de algumas cópias do cassete quando os clones transformantes foram crescidos em meio complexo, indicando baixa estabilidade genética na falta de pressão seletiva. Além disso, dependendo do local de integração, um alto número de cópias mostrou ter um efeito negativo no crescimento da levedura, porém, em nenhum dos casos a produção da proteína heteróloga diminuiu.

## ABSTRACT

Increasing heterologous gene copy number is one of most commonly used strategies to improve *Komagataella phaffii* as an expression system. The use of defective auxotrophic markers is an approach to obtain multicopy integrants. Expression cassettes must be integrated into the *K. phaffii* genome by homologous recombination to ensure mitotic stability. Therefore, it is necessary that the vector carries genomic sequences that will promote integration events. The aim of this work was to identify different sites for the integration of vectors carrying the auxotrophic defective marker *leu2-d* into the genome in order to increase the expression levels of heterologous genes in *K. phaffii*. The targets for integration studied were the following *loci*: 5S rDNA, NTS rDNA and repetitive regions of chromosomes 2 and 3. A vector carrying defective marker *leu2-d* and the EGFP (*Enhanced Green Fluorescent Protein*) reporter gene was constructed. This vector was used as platform to test repetitive sequences as targets for DNA integration. All constructions were used to transform *K. phaffii* M12 (*leu2*). Copy number determination confirmed the generation of multicopy clones with all the constructions. Up to 78 copies of the cassette containing the sequence of chromosome 2 were integrated. All multicopy clones showed more intracellular GFP production compared with a single-copy clone. A 197-fold increase of protein production was observed with the clone containing 78 copies of the cassette.

To validate the strategy presented in this study, the expression of the reporter protein  $\alpha$ -amylase was evaluated. Multicopy clones were obtained with the diverse constructions. Up to 43 copies of the cassette containing the sequence of chromosome 2 were integrated. Amylolytic activity in culture supernatants confirmed amylase production, which was higher for multicopy clones in comparison with a single-copy clone. A 4.6-fold increase of protein production was observed with the clone containing 43 copies of the cassette. Production of the two proteins tested showed the possibility to obtain multicopy clones which produce a larger quantity of the recombinant protein. Nevertheless, loss of some copies of the expression cassette was observed when transformants were grown in complex medium. This indicated low genetic stability in the absence of selective pressure. Moreover, depending of the integration *locus* higher copy number showed a negative effect in yeast growth. However, heterologous protein production was not decreased.

# 1. INTRODUÇÃO

## 1.1 *Komagataella phaffii* e expressão heteróloga

*Komagataella phaffii* (inicialmente classificada como *Pichia pastoris*) é uma levedura metilotrófica, o que significa que é capaz de crescer em metanol como única fonte de carbono e energia. Esta levedura pertence ao grupo dos ascomicetos, é homotética e é haploide a menos que seja submetida a condições de limitação por nitrogênio (Çelik e Çalik, 2012). *P. pastoris* foi isolada em 1920 a partir do exsudato de uma árvore de castanha (Guilliermond, 1920) e foi depositada como a linhagem NRRL-Y1603 (Northern Regional Research Laboratories, Peoria, IL, EUA), CBS 704 (Centraalbureau voor Schimmelcultures, Utrecht, Holanda) ou ATCC 76273 (American Type Culture Collection, EUA). Um isolado de uma árvore de carvalho na Califórnia foi depositado como a linhagem NRRL-Y11430, CBS 7435 ou ATCC 28485.

Após o descobrimento de que *P. pastoris* podia crescer em metanol, houve um grande interesse em utilizar essa levedura na produção de *single-cell protein* (SCP) para alimentação animal devido ao baixo custo da síntese de metanol a partir de gás natural metano. Desta forma, a companhia Phillips Petroleum começou a desenvolver meios de cultura e métodos para crescer *P. pastoris* em metanol. Porém, a crise do petróleo nos anos 70 levou a um aumento drástico do preço do metano, o que tornou economicamente inviável o processo de produção de SCP. Nos anos 80, a Phillips Petroleum em parceria com o Salk Institute Biotechnology/Industrial Associates, Inc. (SIBIA), desenvolveu *P. pastoris* como sistema de expressão heteróloga (Cregg *et al.*, 2000). Atualmente a empresa Invitrogen (Carlsbad, CA, USA) disponibiliza o kit de expressão para seu uso em pesquisa, com diversos tipos de linhagens para usos específicos.

*Pichia pastoris* foi reclassificada no gênero *Komagataella* conforme as análises filogenéticas de quatro genes sequenciados (Kurtzman, 2009). Esse estudo determinou que as linhagens usadas em aplicações biotecnológicas pertencem às espécies *K. phaffii* (linhagens derivadas de CBS 7435) e *K. pastoris* (linhagens derivadas de CBS 704). Porém, o termo “*Pichia pastoris*” continua ainda sendo usado para se referir ao sistema de expressão.

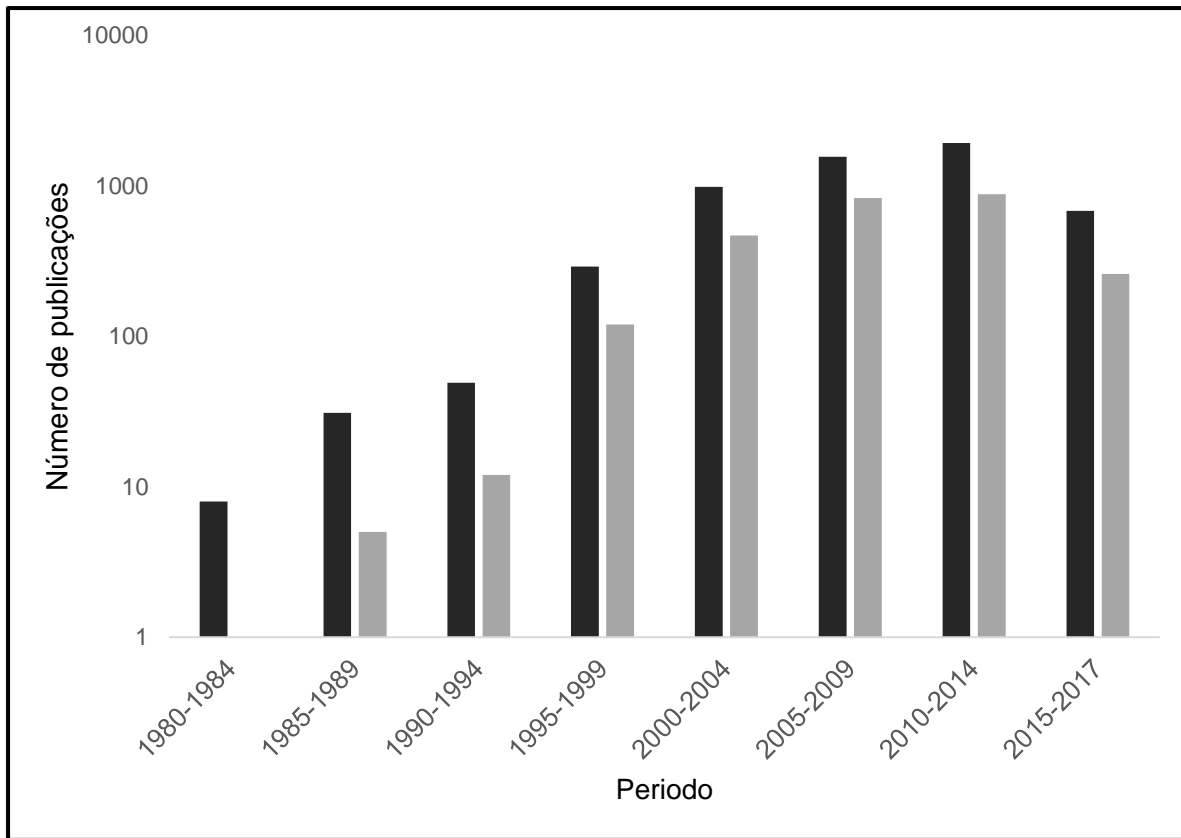
As linhagens comerciais mais usadas GS115 e X-33 (Invitrogen, USA) são derivadas do isolado CBS 7435 e, deste modo, pertencem à espécie *K. phaffii*. A linhagem GS115 é auxotrófica para histidina e é usada como hospedeira para vetores contendo o gene *HIS4* como marca de seleção (Cregg *et al.*, 1985). A linhagem X-33 foi derivada da linhagem GS115 após transformação com o fragmento NaeI-BamHI do vetor pAO815 que contém o gene *HIS4* e, portanto, apresenta fenótipo selvagem (Higgins e Cregg, 1998).

Por ser um organismo unicelular, *K. phaffii* é facilmente cultivável e manipulável. No entanto também é um eucarioto capaz de realizar as modificações pós-traducionais dos eucariotos mais complexos como processamento e dobramento de proteínas, formação de ligações dissulfeto e glicosilação (Cregg *et al.*, 2000). Desta forma, proteínas que ficam dentro de corpos de inclusão inativos quando produzidas em bactéria, podem ser produzidas como moléculas ativamente biológicas em *K. phaffii*. Ademais, o sistema de expressão baseado em *K. phaffii* é mais fácil de usar, cresce mais rápido e é mais econômico do que os sistemas de expressão baseados em células de mamíferos ou insetos (Higgins e Cregg, 1998).

*K. phaffii* não possui a enzima  $\alpha$ -1,3 manosiltransferase, responsável pela formação das ligações  $\alpha$ -1,3-manose em *Saccharomyces cerevisiae*, as quais são altamente imunogênicas. Além disso, a quantidade de resíduos de manose adicionados por *K. phaffii* é menos extensa comparada com *S. cerevisiae* (Trimble *et al.* 1991; Montesino *et al.* 1998; Celik e Calik 2012; Tran *et al.*, 2017). Consequentemente, as modificações pós-traducionais realizadas por *K. phaffii* são mais semelhantes às de células de mamíferos. Outra vantagem desse sistema de expressão é que os níveis de secreção de proteínas endógenas em *K. phaffii* são muito baixos, portanto uma proteína heteróloga secretada constitui a maioria das proteínas totais no meio de cultura, o que facilita a purificação da proteína de interesse produzida neste sistema (Tschopp *et al.* 1987; Cereghino e Cregg 2000). Finalmente, *K. phaffii* desfruta do status GRAS, já que não é patogênica nem toxigênica sendo aprovada pela FDA (*Food and Drug Administration*, EUA) como um aditivo para alimentação animal (FDA, 1993; Ciofalo *et al.*, 2006). Todas estas características fazem de *K. phaffii* um dos mais usados sistemas de produção de proteínas heterólogas (Gasser *et al.*, 2013; Ahmad *et al.*, 2014) assim como uma atraente plataforma para a engenharia de vias metabólicas (Kang *et al.*, 2017).

No gráfico da figura 1, pode ser observado o aumento no número de pesquisas envolvendo esta levedura, desde os anos 80, tanto no seu estudo geral (barras pretas, pesquisa com as palavras '*Pichia pastoris*' no título das publicações), como no seu uso como sistema

de expressão (barras cinzas, pesquisa com as palavras '*Pichia pastoris* expression' no título das publicações).



**Figura 1. Número de publicações usando a levedura *K. phaffii* desde 1980 até o presente ano.** As barras pretas representam as publicações relacionadas com *Pichia pastoris* e as barras cinza representam as publicações referentes à expressão em *Pichia pastoris*. Busca realizada no Google acadêmico no mês de maio de 2017.

Mais de 5000 proteínas já foram produzidas em *K. phaffii* e mais de 70 encontram-se no mercado ou na etapa final de desenvolvimento. A tabela 1 apresenta uma lista com algumas dessas proteínas que foram produzidas usando a plataforma de expressão *Pichia* da RCT (Research Corporation Technologies, Estados Unidos). Biofármacos como Kalbitor® (ecallantide) e VIMPAT® (lacosamida) usados para tratar a angioedema e a epilepsia, respectivamente, são produzidos em *K. phaffii* e foram aprovados pela FDA.

**Tabela 1.** Alguns dos produtos no mercado ou na etapa final de desenvolvimento produzidos em *K. phaffii*\*

<b>Produto</b>	<b>Empresa</b>	<b>Uso</b>
Kalbitor® (Ecallantide: proteína recombinante inibidora de calicreína)	Dyax (Cambridge, MA)	Tratamento do angioedema hereditário
Insugen® (Insulina humana recombinante)	Biocon (Índia)	Terapia da diabetes
Medway (Albumina sérica humana recombinante)	Mitsubishi Tanabe Pharma (Japão)	Expansão do volume sanguíneo
Shanvac™ (Vacina recombinante contra Hepatite B)	Shantha/Sanofi (Índia)	Prevenção da Hepatite B
Shanferon™ (Interferon $\alpha$ 2b recombinante)	Shantha/Sanofi (Índia)	Tratamento da Hepatite C e câncer
Ocriplasmin (Microplasmina recombinante)	ThromboGenics (Bélgica)	Tratamento da adesão vítreo-macular
Nanobody® ALX-0061 (Fragmento recombinante de anticorpo contra o receptor IL-6)	Ablynx (Bélgica)	Tratamento da artrite reumatoide
Nanobody® ALX00171 (Fragmento recombinante de anticorpo anti-RSV)	Ablynx (Bélgica)	Tratamento da infecção pelo vírus sincicial respiratório
HB-EGF (Fator de crescimento similar a EGF de ligação a heparina)	Trillium (Canadá)	Tratamento de cistite intersticial/Tratamento da síndrome da dor da bexiga
Purifine (Fosfolipase C recombinante)	Verenium/DSM (San Diego, CA/Holanda)	Degomagem de fosfolipídios
Tripsina recombinante	Roche Applied Science (Alemanha)	Digestão de proteínas
Colágeno recombinante	Fibrogen (San Francisco, CA)	Reagentes para pesquisa médica/enchimento dérmico
AQUAVAC IPN (Proteínas recombinantes do capsídeo do vírus da necrose pancreática infecciosa)	Merck/Schering Plough Animal Health (Summit, NJ)	Vacinas para a necrose pancreática infecciosa em salmão
Fitase recombinante	Phytex, LLC (Sheridan, IN)	Aditivo para alimentação animal
Nitrato redutase recombinante	The Nitrate Elimination Co. (Lake Linden, MI)	Produtos baseados em enzimas para análise e tratamento de águas
Cistatina C humana recombinante	Scipac (Reino Unido)	Reagente para pesquisa

\*Dados obtidos no site da Research Corporation Technologies ([www.pichia.com](http://www.pichia.com))

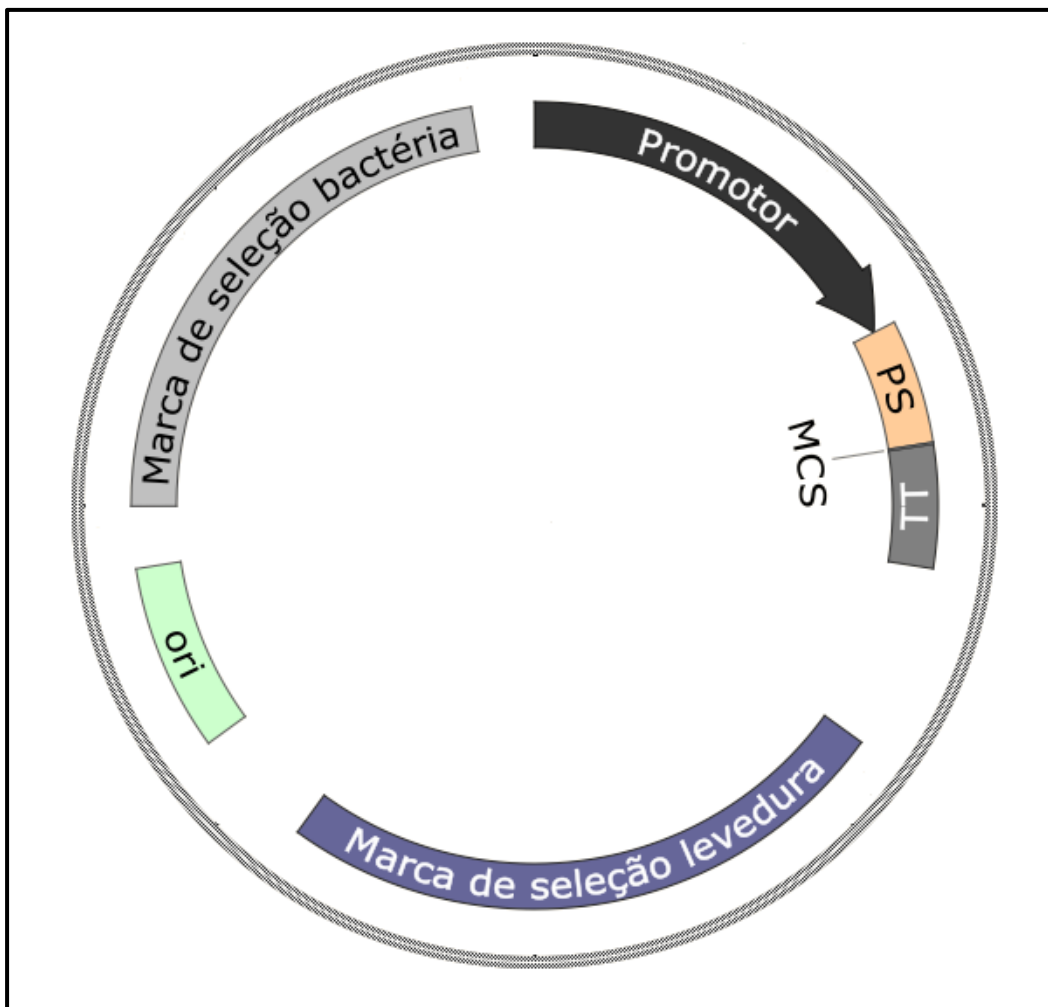
A disponibilidade das sequências do genoma das linhagens selvagens de *K. pastoris* CBS 704 e *K. phaffii* CBS 7435, e da mais usada linhagem mutante derivada GS115 (US Patent 4,879,231, Phillips Petroleum, 1989. ATCC 20864), facilita a sua manipulação genética. Em 2009, De Schutter *et al.* sequenciaram o genoma de *K. phaffii* GS115 e apresentaram uma sequência de 9,43 Mb com uma anotação manualmente curada de 5313 genes codificadores de proteínas. No mesmo ano, Mattanovich *et al.* apresentaram o genoma da linhagem DSMZ 70382 (CBS 704) que apresentou um tamanho de 9,4 Mb. Em 2011, Kuberl *et al.* publicaram o genoma de *K. phaffii* CBS 7435 e apresentaram uma sequência de 9,35 Mb com uma anotação de 5007 genes codificadores de proteínas. Nesse trabalho também foi reportada a sequência do DNA mitocondrial da levedura. Dois trabalhos posteriores foram publicados apresentando uma sequência refinada do genoma da linhagem CBS 7435 (Sturmberger *et al.* 2016; Valli *et al.* 2016). As novas anotações estão disponíveis na base de dados [www.pichiagenome.org](http://www.pichiagenome.org). Em 2016, Love *et al.* realizaram um estudo comparativo dos genomas das linhagens parentais de *K. phaffii* e *K. pastoris* e da linhagem mutante GS115, encontrando vários rearranjos sintênicos entre as duas espécies (blocos de genes realocados no genoma, mas que conservam a ordem e orientação relativa no cromossomo) e 35 mutações não-sinônimas no genoma de GS115. Além disso, foi encontrado um plasmídeo linear de aproximadamente 11 kb presente unicamente na linhagem selvagem de *K. phaffii* contendo sete genes, todos eles homólogos a genes do plasmídeo do sistema *killer* de *Kluyveromyces lactis*. O estudo sugere ainda que o plasmídeo se encontra em baixo número de cópias na linhagem GS115 (Love *et al.*, 2016).

Com relação às ferramentas para manipulação genética, estas se assemelham àquelas usadas em *S. cerevisiae*, um dos sistemas experimentais mais bem caracterizados (Cregg *et al.*, 2000). Recentemente, foi descrito o uso do sistema CRISPR/cas9 em *K. phaffii* como uma ferramenta para realizar modificações genéticas específicas: ruptura gênica e integração de cassetes de DNA homólogo (Weninger *et al.*, 2016).

### **Vetores de expressão**

Os vetores de expressão para *K. phaffii* possuem um cassete de expressão que contém os seguintes elementos: um promotor, seguido de sítios de restrição para a clonagem do gene heterólogo e finalmente, uma sequência terminadora da transcrição que geralmente é o terminador da transcrição do gene *AOX1* da própria *K. phaffii*. As proteínas heterólogas

podem ser produzidas intracelularmente ou secretadas ao meio de cultura. Para a secreção de proteínas, alguns cassetes de expressão contêm sequências sinais localizadas imediatamente depois do promotor e que direcionam a proteína à via de secreção. Os peptídeos sinal mais usados são o da fosfatase ácida de *K. phaffii* ou o fator alfa de *S. cerevisiae*. Outro elemento presente nos vetores de expressão é uma marca de seleção que permite a identificação de clones transformantes. Finalmente, o vetor possui sequências necessárias para a replicação e seleção de transformantes em bactéria durante as etapas de clonagem (Higgins e Cregg, 1998). A figura 2 apresenta o esquema de um vetor de expressão com todos os seus elementos básicos.



**Figura 2. Esquema de um vetor de expressão para *K. phaffii*.** Elementos básicos do vetor de expressão: Promotor, PS: peptídeo sinal, MCS: sítio de clonagem múltipla, TT: terminador da transcrição, marca de seleção para levedura, origem de replicação para bactéria e marca de seleção para bactéria.



## Promotores

Os promotores utilizados nos vetores de expressão para *P. pastoris* podem ser induzíveis ou constitutivos. Os primeiros vetores de expressão de *K. phaffii* foram baseados no promotor do gene *AOX1*, que codifica a enzima álcool oxidase, devido à sua força e por ser altamente regulado pela adição de metanol (Cregg *et al.*, 1987). Porém, o uso do promotor  $P_{AOX1}$  apresenta algumas desvantagens como a dificuldade de monitorar as concentrações de metanol durante a fermentação e o perigo da estocagem do metanol por ser uma substância inflamável. Além disso, o sistema *AOX1* necessita de duas fontes de carbono: glicerol para a fase de crescimento, e metanol para a fase de indução (Macauley-Patrick *et al.*, 2005).

Outros vetores com promotores constitutivos se tornaram disponíveis nos últimos anos. A maioria deles possui o promotor  $P_{GAP}$  que controla a transcrição do gene que codifica a enzima gliceraldeído-3-fosfato desidrogenase (Waterham *et al.*, 1997). No nosso grupo de pesquisa foi isolado o promotor do gene *PGKI* que codifica para 3-fosfoglicerato quinase, enzima glicolítica que transfere um grupo fosforil do 1,3-bifosfoglicerato para ADP, produzindo ATP e 3-fosfoglicerato (De Almeida, Moraes e Torres 2005). Foi observada uma alta eficiência na produção da enzima  $\alpha$ -amilase de *Bacillus subtilis* o que demonstrou a possibilidade de usar o  $P_{PGKI}$  como um promotor constitutivo (Arruda *et al.*, 2016). O promotor  $P_{PGKI}$  também foi usado com sucesso para a produção de quimosina bovina (Araújo, 2008). A grande vantagem de utilizar  $P_{PGKI}$  para controlar a expressão do gene heterólogo é que não é necessário realizar mudanças no meio de cultura para passar de uma fase de crescimento para uma fase de expressão. A fonte de carbono usada durante todo o processo pode ser glicose ou glicerol evitando-se assim os problemas de trabalhar com metanol.

## Marcas de seleção

Independentemente do tipo de vetor de expressão utilizado, este deve possuir uma marca de seleção que permita identificar as células que receberam o DNA exógeno. As marcas de seleção disponíveis para *K. phaffii* podem ser genes de vias biossintéticas como *HIS4*, *ARG4*, *URA3*, dentre outros, ou genes que conferem resistência a drogas como os genes *Sh ble* e *nptII* que conferem resistência a zeocina e G418, respectivamente (Nett e Gerngross 2003; Thor *et al.* 2005). No caso das marcas auxotróficas, são necessárias linhagens hospedeiras nas quais o gene da via biossintética em questão esteja inativo. A maioria das linhagens auxotróficas são obtidas por mutação pontual, porém, a inativação do gene pode ser

feita pela deleção completa ou de uma parte dele como é o caso da linhagem auxotrófica para leucina onde um fragmento de aproximadamente 400 pb do gene *LEU2* (que codifica a enzima  $\beta$ -isopropilmalato desidrogenase) foi deletado (Betancur *et al.*, 2017).

### **Secreção de proteínas heterólogas**

A secreção de proteínas heterólogas geralmente está limitada a proteínas que são secretadas no hospedeiro nativo. A secreção requer a presença de uma sequência sinal para direcionar a proteína à via de secreção. O retículo endoplasmático é a organela onde ocorre a formação de ligações dissulfeto e o dobramento das proteínas antes de serem destinadas ao espaço extracelular. Em condições normais, os peptídeos recém-sintetizados são mantidos no retículo endoplasmático na forma solúvel pela interação com chaperonas como a BiP. Altas quantidades da proteína heteróloga levam à formação de agregados e a uma suprarregulação da resposta a proteínas não-dobradas (UPR, *unfolded protein response*) a qual é uma via de sinalização ativada em resposta ao acúmulo de proteínas não dobradas no retículo endoplasmático. Depois de um sinal prolongado da UPR, é ativada a via de degradação de proteínas do retículo endoplasmático (ERAD, *endoplasmatic reticulum associated protein degradation*) (Mattanovich *et al.*, 2004; Love *et al.*, 2012). Várias estratégias podem ser usadas para superar os gargalos na via secretora, entre elas, as mais usadas são a superprodução de proteínas que ajudam no dobramento, como BiP/Kar2p, DnaJ, PDI, PPIs e Ero1p, ou a superexpressão do gene *HAC1*, que é um regulador da transcrição de genes da via da UPR (Ahmad *et al.*, 2014).

O amplo uso de *K. phaffii* para a produção de proteínas de interesse biotecnológico faz necessária a busca de novas estratégias que permitam o melhoramento genético de linhagens a fim de otimizar os níveis de expressão de genes heterólogos. Uma abordagem bem estabelecida para se aumentar esta produção é garantir um nível suficiente de transcrição do gene heterólogo, o que pode ser obtido com o uso de promotores fortes, clones contendo um número maior de cópias do gene, ou a combinação de ambos.

## 1.2 Integração múltipla

Para garantir estabilidade genética, os vetores e cassetes de expressão devem ser integrados no genoma de *K. phaffii*. Geralmente, o vetor de expressão é linearizado em uma das regiões que tem identidade com sequências homólogas no genoma da levedura (promotor, marca auxotrófica ou fragmento 3' do gene *AOX1*). A recombinação homóloga das extremidades linearizadas do vetor resulta em um tipo de integração conhecida como *single-crossover* e, dessa forma, a maioria dos clones transformantes possui apenas uma cópia integrada do vetor de expressão. Porém, eventos de integração múltipla no mesmo *locus* ocorrem espontaneamente com baixa frequência (Cregg *et al.*, 2000).

Os altos níveis da enzima álcool oxidase produzida pelo gene nativo *AOX1* em *K. phaffii* (5-30% das proteínas totais após indução com metanol), sugeriram que uma única cópia do gene heterólogo sob controle do promotor do gene *AOX1* seria suficiente para a produção da proteína de interesse. Alguns estudos apoiaram essa hipótese, como a produção da enzima invertase (Tschopp *et al.*, 1987) ou do antígeno de superfície da hepatite B (HBsAg) (Cregg *et al.*, 1987) que foram produzidos em altos níveis com uma única cópia do vetor integrada no genoma de *K. phaffii*.

Contudo, alguns trabalhos subsequentes mostraram níveis de expressão muito baixos ao usar transformantes contendo uma única cópia do gene heterólogo e o uso de clones multicópia resultou em um aumento da produção da proteína de interesse (Clare *et al.*, 1991a; Clare *et al.*, 1991b; Vassileva *et al.*, 2001; Mansur *et al.*, 2005; Marx *et al.*, 2009). Porém, a relação entre número de cópias do gene heterólogo e o produto não é sempre linear e, em alguns casos, o aumento no número de cópias pode levar a uma diminuição da proteína heteróloga, principalmente quando esta é secretada (Hohenblum *et al.*, 2004; Zhu *et al.*, 2009).

No caso do fragmento C da toxina do tétano, foi observado que o número de cópias é o fator mais importante que afeta a produção de proteínas intracelulares em *K. phaffii* (Clare *et al.*, 1991a). Nesse estudo, o aumento do número de cópias até 14, resultou em um incremento de 10% na produção da proteína. O local de integração, assim como o tipo de integração e o fenótipo de utilização de metanol mostraram um efeito menor nos níveis de produção dessa toxina. Porém, a expressão foi maior nos clones com fenótipo Mut<sup>S</sup>, devido a que o

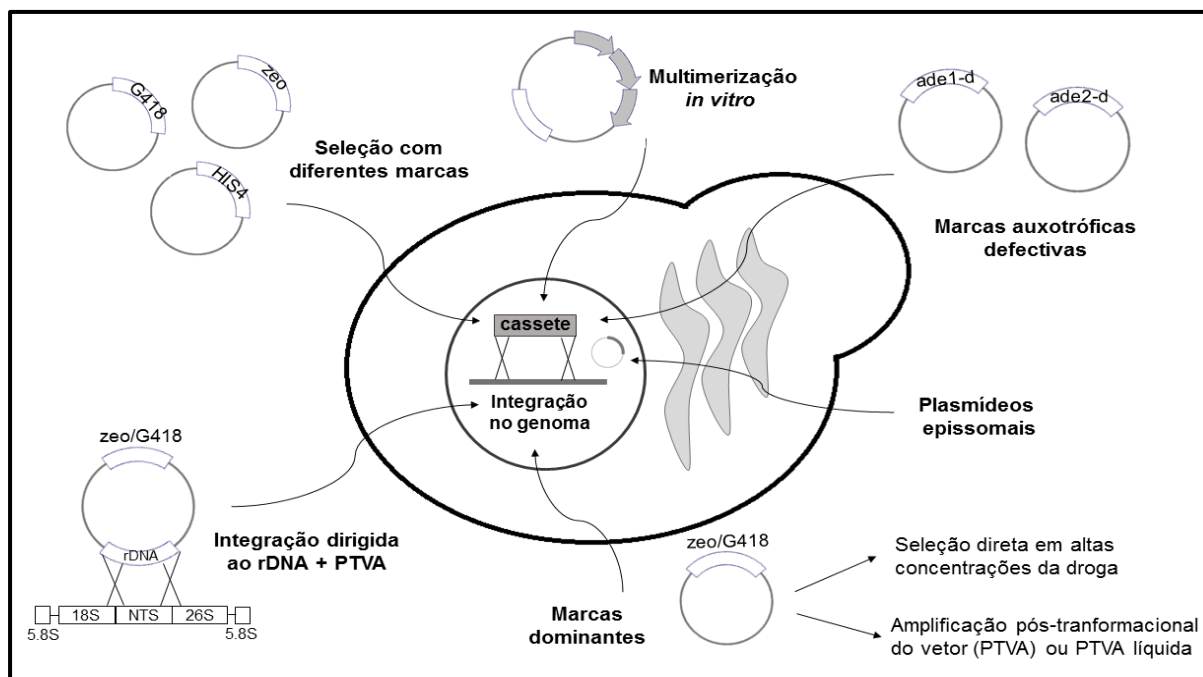
crescimento mais lento em metanol deixa disponíveis uma maior quantidade de substratos e precursores para a produção da proteína heteróloga.

O aumento do número de cópias do gene até 19 cópias levou a um aumento proporcional na secreção do fator de crescimento epidermal murino (Clare *et al.*, 1991b). Uma correlação linear entre o número de cópias e a produção intracelular do antígeno de superfície da hepatite B foi observada quando o gene foi expresso sob o controle do  $P_{GAP}$  (Vassileva *et al.*, 2001). A produção extracelular de miniproinsulina (MPI) também mostrou uma correlação direta ao aumentar o número de cópias do gene (Mansur *et al.*, 2005).

A integração múltipla no *locus* do rDNA apresentou resultados diferentes na produção de uma proteína intracelular e uma proteína secretada. A produção intracelular de superóxido dismutase humana (hSOD) em *K. phaffii* apresentou correlação direta com o número de cópias enquanto que a secreção de albumina de soro humana (HSA) em *K. pastoris* mostrou um aumento até aproximadamente 7 cópias, com uma diminuição da produção com números de cópias maiores (Marx *et al.*, 2009).

A secreção de tripsinogênio humano sob controle do promotor *AOX1* em *K. phaffii*, diminuiu quando três cópias do cassete de expressão foram integradas e uma grande parte da proteína foi retida na fração insolúvel do retículo endoplasmático (Hohenblum *et al.*, 2004). O incremento até 12 cópias do gene PIP teve um efeito positivo na produção do precursor da insulina porcina, porém, qualquer aumento adicional no número de cópias resultou na redução da produção da proteína, assim como na redução do crescimento da levedura (Zhu *et al.*, 2009).

Várias estratégias podem ser empregadas para selecionar clones com maior número de cópias do vetor integradas no genoma. Duas revisões recentes descrevem estas estratégias (Aw e Polizzi 2013; Piva *et al.*, 2017) que são resumidamente mostradas na figura 3.



**Figura 3. Estratégias para gerar clones multicópia em *K. phaffii*.** Clones multicópia podem ser obtidos usando diferentes estratégias: multimerização *in vitro*, transformação sequencial com vetores contendo diferentes marcas de seleção, seleção de colônias sob pressão seletiva usando marcas dominantes (resistência a drogas), amplificação com o método PTVA (ou PTVA líquida), integração dirigida ao *locus* do rDNA, uso de vetores epissomais ou uso de marcas auxotróficas defectivas. Adaptado de Piva *et al.* (2017).

A levedura pode ser transformada com vetores contendo várias cópias do gene clonado *in tandem* (multimerização *in vitro*) (Sreekrishna e Kropp 1996; Li *et al.* 2009) ou podem ser realizadas repetidas rodadas de transformação usando diferentes marcas de seleção, porém, neste caso, as etapas de clonagem são laboriosas e o número de cópias é limitado. Para a multimerização *in vitro*, geralmente é usado o vetor pAO815 (Invitrogen, EUA) que contém um sítio para BglII *upstream* ao promotor  $P_{AOX1}$  e um único sítio para BamHI *downstream* ao terminador da transcrição do gene *AOX1* (3' *AOX1* TT). As seguintes etapas são necessárias para gerar um vetor contendo múltiplas cópias do gene de interesse: 1) o gene é clonado no único sítio para EcoRI no vetor (após o promotor  $P_{AOX1}$ ); 2) a construção é digerida com BglII e BamHI para liberar o cassete de expressão contendo o promotor  $P_{AOX1}$ , o gene de interesse e o 3' *AOX1* TT; 3) várias cópias do cassete de expressão são ligadas *in vitro*; 4) as múltiplas cópias são inseridas no vetor pAO815 linearizado com BamHI.

Outra opção consiste no uso de marcas dominantes que conferem resistência a antibióticos. Nesse caso, buscam-se clones transformantes que cresçam em concentrações

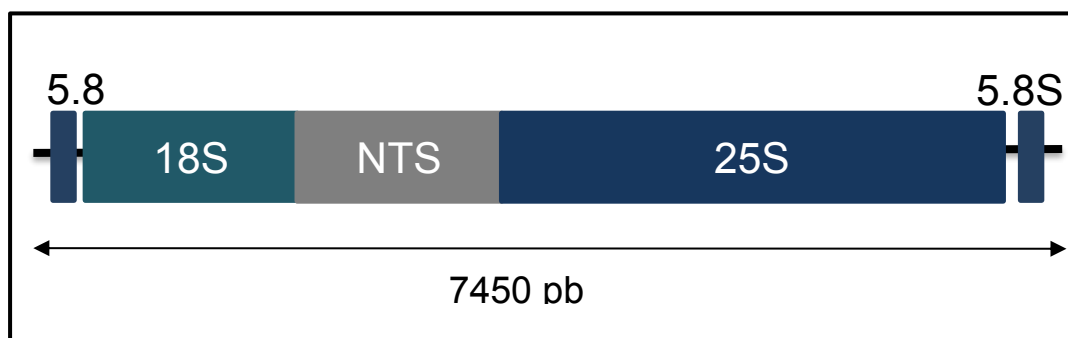
superiores do antibiótico usado. Um estudo prévio mostrou que a seleção dessa forma permitiu isolar integrantes multicópia esporádicos o que resultou em um aumento na produção da proteína de interesse (Chen *et al.*, 2006). As marcas dominantes também permitem isolar clones que tenham integrado múltiplas cópias pelo método PTVA (*post-transformational vector amplification*) proposto por Sunga *et al.* em 2008. Usando vetores com as marcas que conferem resistência a zeocina ou a G418, foi demonstrado que células de *K. phaffii* transformadas com uma ou poucas cópias do vetor podem ser selecionadas em concentrações maiores do antibiótico muito tempo depois da transformação. As células transformantes são crescidas em placas com concentrações crescentes do antibiótico com aproximadamente cinco dias de incubação entre cada etapa. Durante esta fase de crescimento, o número de cópias do gene de resistência é aumentado para que a célula possa se adaptar a uma concentração do antibiótico maior. Foi demonstrado que nessas condições a célula amplifica todo o cassete ou vetor de expressão incluindo o gene de interesse (Sunga, Tolstorukov e Cregg 2008). Recentemente, foi descrita uma variação deste método: a PTVA líquida (Aw e Polizzi 2016). Esta é uma alternativa mais rápida e econômica comparada com a PTVA tradicional já que os tempos de incubação são reduzidos a 12 ou 24 h e os volumes de meio são menores o que diminui a quantidade de antibiótico requerido. Além dos altos preços das drogas (especialmente zeocina), a desvantagem de usar essas marcas dominantes é que um número significativo de clones apresenta aumento na resistência por razões desconhecidas e não por um aumento no número de cópias do vetor. Além disso, existe o risco de que organismos geneticamente modificados sejam liberados acidentalmente no ambiente em processos em larga escala. Isto poderia levar a uma transferência horizontal do gene de resistência a organismos presentes no ambiente (Dröge, Pühler e Selbitschka 1998).

Uma estratégia alternativa está baseada no uso de marcas auxotróficas defectivas que, tipicamente, representam genes fracamente transcritos, já que apresentam sequências promotoras truncadas. Para compensar esses baixos níveis de expressão a célula necessita um alto número de cópias dessa marca a fim de assegurar a prototrofia e, conseqüentemente, também será aumentado o número de cópias do gene heterólogo (Kazemi Seresht *et al.*, 2013).

Recentemente, um plasmídeo epissomal contendo a sequência panARS foi testado para a expressão heteróloga em *K. phaffii* (Camattari *et al.*, 2016). A sequência panARS de replicação autônoma de *K. lactis* conferiu estabilidade ao plasmídeo e permitiu selecionar

clones com entre 6 e 19 cópias do mesmo. O uso deste plasmídeo epissomal demonstrou ser uma abordagem adequada para incrementar o número de cópias do gene de interesse com uma alta eficiência de transformação e uma maior homogeneidade entre os clones transformantes.

Dependendo da forma como o vetor se integra no genoma da levedura é possível também obter clones multicópia. Existem diversas maneiras de direcionar a integração do vetor com esse fim. Uma delas é a chamada integração *in tandem* por adição na qual várias cópias do vetor são integradas por recombinação homóloga no mesmo *locus*. A desvantagem desta estratégia é que podem haver eventos de recombinação homóloga intramolecular entre as sequências integradas o que pode resultar na eliminação da construção na ausência de pressão seletiva. Outra opção é dirigir o gene ou cassete de expressão para integração no *cluster* do DNA ribossomal (rDNA) que se organiza em unidades *in tandem* (figura 4). No genoma de *K. phaffii* GS115 foi identificado um *contig* separado de 7450 pb correspondente ao rDNA presente nos quatro cromossomos e que está repetido aproximadamente 16 vezes (De Schutter *et al.*, 2009). Porém, um estudo mais recente demonstrou que o *cluster* do rDNA está localizado numa posição subtelomérica nos cromossomos 1, 3 e 4 nas linhagens nativas de *K. phaffii* e *K. pastoris*, enquanto que na linhagem GS115 o mesmo está localizado unicamente no cromossomo 1 (Love *et al.*, 2016). O mecanismo de integração no rDNA consiste em integrar primeiro um baixo número de cópias nesse *locus* seguido de uma amplificação por forte pressão seletiva (Lopes *et al.*, 1991). Neste caso, eventos de recombinação intramolecular também podem eliminar a construção. O uso do *locus* do rDNA, mais especificamente a região espaçadora não transcrita (NTS), como alvo de integração em *K. phaffii* e em *K. pastoris* em combinação com a amplificação pós-transformacional do vetor permitiram obter clones multicópia (Marx *et al.*, 2009).



**Figura 4.** Representação esquemática do *locus* do rDNA em *K. phaffii*. As unidades 5.8S, 18S e 25S são codificados *in tandem* separados por um espaçador não transcrito, NTS.

Uma terceira alternativa é dirigir a integração para sequências repetidas não essenciais que se encontram espalhadas no genoma, como as sequências solo LTRs (*solo long terminal repeats*) e Ty de *S. cerevisiae* (Wang *et al.*, 1996; Flagfeldt *et al.*, 2009). Além destas sequências, o rDNA 5S é também um alvo para integração em *K. phaffii* já que, ao contrário do que se observa em *S. cerevisiae*, esta sequência não faz parte do cluster do rDNA. Em *S. cerevisiae* o rDNA 5S encontra-se no *locus* do rDNA, porém, é transcrito pela RNA polimerase III separadamente e na orientação contrária à das unidades 25S, 5.8S e 18S que são transcritas pela RNA polimerase I na forma de um único precursor. Na montagem do genoma de *K. phaffii* GS115 foram identificadas 21 cópias do rDNA 5S distribuídas ao longo dos quatro cromossomos (De Schutter *et al.* 2009; Love *et al.* 2016). A estratégia de integração dispersa resulta em transformantes geneticamente mais estáveis já que a probabilidade de recombinação homóloga entre as sequências integradas é mínima.

### 1.3 Marcas auxotróficas defectivas

As marcas auxotróficas defectivas consistem em genes de vias biossintéticas que possuem seu promotor truncado. Isto faz com que a taxa de transcrição do gene seja baixa e para compensar isto a levedura tem que integrar um alto número de cópias da marca para poder crescer em meio mínimo. Um exemplo é a marca *leu2-d*, alelo truncado do gene *LEU2*, que inicialmente foi usada em um plasmídeo para *S. cerevisiae* o qual foi mantido em alto número de cópias sob pressão seletiva (Erhart e Hollenberg 1983). O alelo *leu2-d* possui apenas 29 pb do promotor original e seu uso como marca de seleção, em combinação com o uso de meio complexo, propiciou um aumento da produção de ácido artemisínico em 8,1 vezes em *S. cerevisiae* (Ro *et al.*, 2008). Esses resultados demonstraram que a marca defectiva permite a levedura manter o plasmídeo com alta estabilidade em meio não seletivo. A marca defectiva *leu2-d* já foi usada com sucesso em *S. cerevisiae* para aumentar a estabilidade do plasmídeo e a produção de várias proteínas como o inibidor da proteinase  $\alpha$ -1 humano (Hoylaerts *et al.*, 1986) ou a toxina *killer* K2 (Servienė e Melvydas 2001).

Outras marcas defectivas testadas em *S. cerevisiae* são os alelos *ura3-d*, *trp1-d* e *his3-d*. A diminuição gradual do tamanho do promotor do gene *URA3* levou a um aumento no número de cópias do plasmídeo e no rendimento de um precursor análogo da insulina (Kazemi Seresht *et al.*, 2013). A marca *ura3-d* também foi usada na levedura *Yarrowia*



*lipolytica* para aumentar o número de cópias, o que levou a um aumento da produção de lipase de 40 vezes em comparação com a linhagem contendo uma única cópia do gene *LIP2* (Pignede *et al.*, 2000). A marca defectiva *trp1-d* foi usada em *S. cerevisiae* para gerar clones multicópia produtores de  $\alpha$ -amilase (Nieto *et al.*, 1999). Cassetes de expressão contendo as marcas *ura3-d*, *leu2-d*, *his3-d* e *trp1-d* foram usados para obter linhagens de *S. cerevisiae* multicópia e aumentar a produção da proteína do capsídeo do vírus da necrose da garoupa vermelha (RG-NNVCP) e a produção de esqualeno (Moon *et al.*, 2016).

Em *K. phaffii* o uso dos genes *ADE1* e *ADE2* com promotores truncados também favoreceu a integração múltipla de plasmídeos e aumento na produção de proteínas heterólogas (Du, Battles e Nett 2012).

Comparado ao uso de marcas dominantes, a utilização de marcas auxotróficas permite diminuir o custo do processo de produção de proteínas recombinantes ao prescindir do uso de antibióticos e evita o risco da transferência horizontal dos genes que conferem resistência.

Em um trabalho anterior realizado no laboratório de Biotecnologia de Leveduras da UnB foi construído um clone de *K. phaffii* mutante para o gene *LEU2* (Betancur *et al.*, 2017) que pode ser utilizado como plataforma para se avaliar o uso da marca *leu2-d*. Contudo, é necessário buscar *loci* genômicos que propiciem integração estável e multicópia do transgene para que esse sistema possa ser implementado em escala industrial. No presente trabalho, duas estratégias de integração múltipla foram avaliadas: 1) integração *in tandem* nos *loci* *PGK1*, NTS rDNA e em outras regiões repetitivas identificadas por análise bioinformática; 2) integração dispersa nos *loci* rDNA 5S. A combinação do uso da marca *leu2-d* e a identificação de novos alvos de integração são abordagens inovadoras ainda não testadas em *K. phaffii* para aumentar os níveis de expressão de transgenes nesse sistema.

## 2. OBJETIVOS

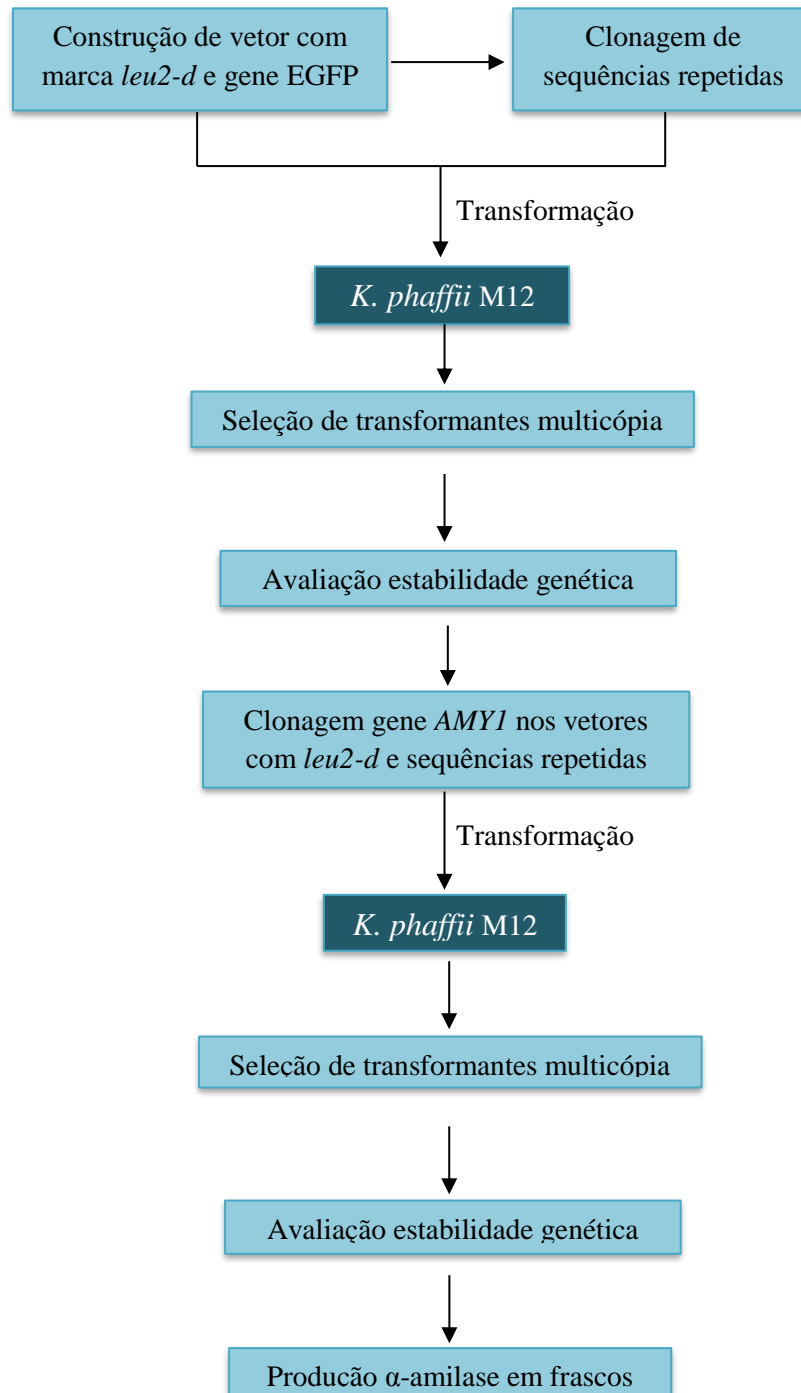
### 2.1 *Objetivo geral*

Avaliar diferentes *loci* para promover a integração múltipla de cassetes de expressão no genoma de *K. phaffii*.

### 2.2 *Metas*

- Construir um vetor de expressão baseado na marca auxotrófica *leu2-d* contendo o gene repórter EGFP sob controle do promotor  $P_{PGK1}$ .
- Construir vetores de expressão contendo a marca *leu2-d*, o gene repórter EGFP e sequências repetidas do genoma de *K. phaffii*.
- Construir vetores de expressão contendo a marca *leu2-d*, sequências repetidas do genoma de *K. phaffii* e o gene *AMY1* fusionado ao peptídeo sinal do fator  $\alpha$ .
- Transformar linhagem auxotrófica M12 de *K. phaffii* (*leu2*) com as diferentes construções e determinar o número de cópias do vetor integradas no genoma.
- Avaliar a produção de GFP e de  $\alpha$ -amilase dos diferentes transformantes e sua estabilidade genética.

### 3. ESTRATÉGIA



## 4. MATERIAL E MÉTODOS

### 4.1 Material

#### 4.1.1 Células

##### Bactérias:

- *Escherichia coli* XL10-gold (Agilent Technologies, EUA):  $Tet^R \Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacI^qZ\Delta M15 Tn10 (Tet^R) Amy Cam^R]$
- *Escherichia coli* HST08 Stellar Competent Cells (Clontech EUA):  $F^-$ ,  $endA1$ ,  $supE44$ ,  $thi-1$ ,  $recA1$ ,  $relA1$ ,  $gyrA96$ ,  $phoA$ ,  $\Phi80d lacZ\Delta M15$ ,  $\Delta (lacZYA^- argF) U169$ ,  $\Delta (mrr^- hsdRMS^- mcrBC)$ ,  $\Delta mcrA$ ,  $\lambda^-$

##### Leveduras:

- *K. phaffii* X-33:  $his4::HIS4$  (derivada da linhagem GS115) (Invitrogen, EUA)
- *K. phaffii* M12:  $leu2$  (derivada da linhagem X-33) (Betancur *et al.*, 2017)

#### 4.1.2 Meios de cultura

##### LB

Extrato de levedura	0,5% (p/v)
Peptona	1% (p/v)
NaCl	1% (p/v)

O pH foi ajustado para 7,2. Para o meio sólido foi adicionado ágar 1,5% (p/v).

## **SOB**

Triptona	20 g/L
Extrato de levedura	5 g/L
NaCl	0,6 g/L
KCl	0,5 g/L
MgCl <sub>2</sub>	10 mM
MgSO <sub>4</sub>	10 mM

Soluções de MgCl<sub>2</sub> e MgSO<sub>4</sub> na concentração de 1 M foram filtradas e adicionadas ao meio autoclavado para a concentração final de 10 mM.

## **YPD**

Extrato de levedura	1% (p/v)
Peptona	2% (p/v)
Glicose	2% (p/v)

Para o meio sólido foi adicionado ágar 2% (p/v).

## **MD (Meio Mínimo com Dextrose)**

YNB com sulfato de amônio	1,34% (p/v)
Biotina	4x10 <sup>-5</sup> % (p/v)
Glicose	2% (p/v)
Ágar	1,5% (p/v)

## **BMD (Meio MD tamponado)**

YNB com sulfato de amônio	1,34% (p/v)
Biotina	4x10 <sup>-5</sup> % (p/v)
Glicose	2% (p/v)
Fosfato de potássio pH 6,0	100 mM
Ágar	1,5% (p/v)

### **BMDA (Meio Mínimo com Dextrose e Amido tamponado)**

YNB com sulfato de amônio	1,34% (p/v)
Biotina	$4 \times 10^{-5}$ % (p/v)
Glicose	2% (p/v)
Tampão Asp-Glu pH 5,5	0,4% (p/v)
Amido	1% (p/v)
Ágar	0,59% (p/v)

Para os meios mínimos foram feitas soluções estoque de todos os reagentes e misturadas assepticamente no momento de preparar as placas. A solução de glicose 10X, a solução de fosfato de potássio 1 M pH 6,0 e a solução de ágar 2% (p/v) foram esterilizadas por autoclavagem separadamente. A solução de biotina 500X, a solução YNB com sulfato de amônio 10X e o tampão Asp-Glu 20X foram esterilizadas por filtração.

### **4.1.3 Soluções**

#### **Soluções estoque para meios de cultura**

- Solução YNB (Yeast Nitrogen Base) com sulfato de amônio 10X

Yeast Nitrogen Base w/o Amino acids (Difco) 3,4% (p/v)

$(\text{NH}_4)_2\text{SO}_4$  10% (p/v)

Esterilizada por filtração.

- Solução de glicose 10X

Glicose 20% (p/v)

- Solução de biotina 500X

Biotina 0,02% (p/v)

Esterilizada por filtração.

- Tampão Asp-Glu

Ácido aspártico      8% (p/v)

Ácido glutâmico      8% (p/v)

Esterilizado por filtração.

## **Antibióticos**

- Solução de ampicilina (1000X)

Ampicilina              100 mg/mL

Dissolvida em água milli Q e esterilizada por filtração.

- Solução de kanamicina (1000X)

Kanamicina              50 mg/mL

Dissolvida em água milli Q e esterilizada por filtração.

## **Soluções para células bacterianas competentes**

- Tampão de transformação I

RbCl                      12 g/L

MnCl<sub>2</sub>·4H<sub>2</sub>O              9,9 g/L

Acetato de potássio      0,03 M

CaCl<sub>2</sub>·2H<sub>2</sub>O              1,5 g/L

Glicerol                 150 g/L

Uma solução estoque de acetato de potássio 1 M foi preparada e seu pH ajustado para 7,5 com ácido acético glacial. A solução foi esterilizada por filtração. O pH do tampão de transformação foi ajustado para 5,8 com ácido acético 0,2 M e a solução foi esterilizada por filtração.

- Tampão de transformação II

MOPS	0,02 M
RbCl	1,2 g/L
CaCl <sub>2</sub>	11 g/L
Glicerol	150 g/L

Uma solução estoque de MOPS 1 M foi preparada e seu pH ajustado para 6,8 com NaOH. A solução foi esterilizada por filtração. O pH do tampão de transformação foi ajustado para 6,8 com NaOH e a solução foi esterilizada por filtração.

### **Soluções para preparação de DNA plasmidial (miniprep)**

- Solução I

Tris-HCl pH 8,0	25 mM
EDTA pH 8,0	10 mM

- Solução II

NaOH	0,2 M
SDS	1% (p/v)

Solução preparada no momento de uso.

- Solução III

Acetato de sódio	3 M
Ácido acético	2 M

O pH foi ajustado para 5,0.





- SSC 20X

Citrato de sódio 1 M

NaCl 5 M

### Soluções para eletroforese em campo pulsado

- Solução de esferoplastos

$\beta$ -Mercaptoetanol 7,5% (v/v)

EDTA pH 8,0 0,5 M

Liticase 0,1 mg/mL

- Solução de protease

Sarcosil 1,5% (p/v)

EDTA pH 9,2 0,5 M

Proteinase K 1 mg/mL

- Tampão Tris-borato EDTA (TBE) 10X

Tris 0,89 M

EDTA pH 8,0 0,02 M

Ácido bórico 0,89 M

### Soluções para eletroforese em gel de poliacrilamida

- Acrilamida:bisacrilamida (29:1)

Acrilamida 29% (p/v)

Bisacrilamida 1% (p/v)

- Tris-HCl 1,5 M pH 8,8

- Tris-HCl 1 M pH 6,8

- SDS 10% (p/v)
- Persulfato de amônio 10% (p/v)

Armazenado a -20°C.

- Tampão de amostra não desnaturante 5X

Tris-HCl pH 6,8	60 mM
Glicerol	25% (v/v)
Azul de bromofenol	0,1% (p/v)

- Tampão de amostra desnaturante 5X

Tris-HCl pH 6,8	60 mM
SDS	2% (p/v)
β-Mercaptoetanol	14,4 mM
Glicerol	25% (v/v)
Azul de bromofenol	0,1% (p/v)

- Tampão de corrida Tris-Glicina 5X

Tris	16,7 g/L
Glicina	104,5 g/L
SDS	0,5% (p/v)

- Solução corante

Coomassie brilliant blue G-250	0,25% (p/v)
Metanol	30% (v/v)
Ácido acético glacial	7% (v/v)

- Solução descorante

Metanol	30% (v/v)
Ácido acético glacial	7% (v/v)

## **Soluções para extração de DNA genômico de levedura**

- Tampão SE

Sorbitol	0,9 M
EDTA pH 7,5	100 mM

- Tampão TE<sub>20</sub>

Tris-HCl pH 7,5	10 mM
EDTA pH 7,5	20 mM

- Clorofane

Fenol	25% (v/v)
Clorofórmio	24% (v/v)
Álcool isoamílico	1% (v/v)

## **Soluções para citometria de fluxo**

- PBS

NaCl	137 mM
KCl	2,7 mM
Na <sub>2</sub> HPO <sub>4</sub>	10 mM
KH <sub>2</sub> PO <sub>4</sub>	2 mM

## **Soluções para ensaio enzimático de degradação de amido**

- Tampão acetato de sódio

Acetato de sódio 0,5 M

O pH foi ajustado para 6,0 com ácido acético.

- Solução de amido 0,5% (p/v)

Preparada em água destilada e esquentada em forno micro-ondas até dissolver, sem ferver.

- Solução de iodo 1% (p/v)

Dissolvido em etanol.

- Solução de iodeto de potássio (KI) 10% (p/v)

- Ácido acético 1M

- Solução FUWA (Fuwa, 1954)

Solução de iodo 1%	1 parte
Solução de KI 10%	1 parte
Água destilada	3 partes

### **Soluções para zimograma**

- Solução reveladora

Iodo	1 mM
Iodeto de potássio	0,5 M

#### **4.1.4 Kits**

- Qiagen Plasmid Maxi kit (Qiagen): purificação de plasmídeos em grande escala.
- In-Fusion HD Cloning Kit (Clontech): clonagem de fragmentos por recombinação homóloga.
- Wizard SV Gel and PCR Clean-Up System (Promega): purificação de fragmentos de DNA a partir de gel de agarose e purificação de produtos de PCR.

- AlkPhos Direct Labeling and CDP-Star Detection System (GE Life Sciences):  
marcação de sonda e detecção para análise por Southern blot.

#### 4.1.5 Oligonucleotídeos (*primers*)

A tabela 2 apresenta os *primers* utilizados neste trabalho com seus respectivos sítios de restrição sublinhados.

**Tabela 2.** *Primers* utilizados neste trabalho\*

<i>Primers</i>	Sequência (5' → 3')	Sítio de restrição
5-leud	G <u>GATCT</u> ATATATATTTCAAGGATATACCATTCTAATG	BglII
3-leud	G <u>GATCT</u> GTTTCATGATTTTCTGTTACACC	BglII
PGK-chr2	gaagcaactctggc <u>AGATCT</u> GCGAGGCAAGCATCTACTA	BglII
leud-chr2	ctcagttggctcggcAACATTTTGAAGCTATGGTGTGTGGG	
PGK-chr3	cgcttgccggccggc <u>AGATCT</u> GCGAGGCAAGCATCTACTA	BglII
leud-chr3	tgccgcgaggtggcAACATTTTGAAGCTATGGTGTGTGGG	
PGK-5S	gtgctgcaactctggc <u>AGATCT</u> GCGAGGCAAGCATCTACTA	BglII
leud-5S	atggccgcaaccggcAACATTTTGAAGCTATGGTGTGTGGG	
PGK-NTS	tccaaagattactag <u>AGATCT</u> GCGAGGCAAGCATCTACTA	BglII
leud-NTS	tgaaaagtaactagAACATTTTGAAGCTATGGTGTGTGGG	
18SF-pPCV	ggttccggtggagatc <u>GTTTAAACT</u> ACCATCGAAAGTTGA	EcoRV e PmeI
25SR-pPCV	actagagtggatagatc <u>GTTTAAACT</u> TAGACCGTCGTGAGACAGG	EcoRV e PmeI
EGFP-Eco	<u>GGAATTC</u> ATGGTGAGCAAGGGCGAG	EcoRI
EGFP-Not	<u>GGCGGCCGCTT</u> ACTTGTACAGCTCGTCCATG	NotI
Falphaif-F	gattacgaaaggatccATGAGATTCCCATCTATCTTCACTG	BamHI
AMYif-R	atggtcgacgcccgcCTAGGAGGACCACGTAAACTCGATAT	NotI

\*Em caixa baixa, as regiões de homologia para realizar a clonagem usando o sistema In-Fusion. Os sítios de restrição estão sublinhados.

#### 4.1.6 Marcadores de massa molecular para DNA

- 2-log DNA ladder, 100-10.000 pb (New England Biolabs).
- O'GeneRuler 1 kb Plus DNA Ladder 75-20.000 pb (Thermo Scientific).
- 1 kb DNA Ladder, 50-10.000 pb (New England Biolabs).

#### 4.1.7 Enzimas de restrição

A tabela 3 apresenta as enzimas de restrição usadas neste trabalho indicando o sítio de clivagem, assim como a temperatura de incubação e o tampão recomendado para cada enzima.

**Tabela 3.** Enzimas de restrição usadas

Enzima	Sítio de clivagem	Tampão	Temperatura de incubação
BamHI	G <sup>+</sup> GATCC	NEBuffer 3.1	37 °C
BglII	A <sup>+</sup> GATCT	NEBuffer 3.1	37 °C
EcoRI	G <sup>+</sup> AATTC	NEBuffer EcoRI	37 °C
EcoRV	GAT <sup>+</sup> ATC	NEBuffer 3.1	37 °C
KpnI	GGTAC <sup>+</sup> C	NEBuffer 1.1	37 °C
NotI	GC <sup>+</sup> GGCCGC	NEBuffer 3.1	37 °C
PmeI	GTTT <sup>+</sup> AAAC	CutSmart	37 °C
SacI	GAGCT <sup>+</sup> C	NEBuffer 1.1	37 °C
SfoI	GGC <sup>+</sup> GCC	CutSmart	37 °C
SpeI	A <sup>+</sup> CTAGT	CutSmart	37 °C

#### 4.1.8 Vetores

A tabela 4 exibe todos os vetores usados neste trabalho e apresenta uma breve descrição de cada um deles.

**Tabela 4.** Vetores utilizados

Vetor	Descrição
pPICK2	Vetor de expressão para <i>K. phaffii</i> construído em nosso laboratório (Arruda <i>et al.</i> , 2016) derivado do vetor comercial pPICZαA (Invitrogen, EUA). Contém o cassete de expressão do gene <i>kan<sup>R</sup></i> e o promotor <i>PGK1</i> .
pKld	Vetor construído neste trabalho derivado do pPICK2, contendo o alelo <i>leu2-d</i> como marca de seleção.
pPICK-GFP	Vetor de expressão para <i>K. phaffii</i> construído em nosso laboratório para expressar e secretar proteína verde fluorescente. Este é derivado do vetor comercial pPIC9 (Invitrogen, EUA).
pKld-GFP	Vetor construído neste trabalho para a expressão de GFP contendo o alelo <i>leu2-d</i> como marca de seleção.
pKL2-GFP	Vetor construído no nosso laboratório para a expressão de GFP contendo o gene <i>LEU2</i> como marca de seleção.
pPCV	Vetor de clonagem construído em nosso laboratório (Janner <i>et al.</i> , 2013) derivado do vetor comercial pBlueScript II KS (Agilent Technologies, EUA).
pBS-Chr2	Vetor construído neste trabalho derivado do vetor comercial pBlueScript II KS (Agilent Technologies, EUA), contendo uma sequência repetida do cromossomo 2 de <i>K. phaffii</i> .
pBS-Chr3	Vetor construído neste trabalho derivado do vetor comercial pBlueScript II KS (Agilent Technologies, EUA), contendo uma sequência repetida do cromossomo 3 de <i>K. phaffii</i> .
pBS-5S	Vetor construído neste trabalho derivado do vetor comercial pBlueScript II KS (Agilent Technologies, EUA), contendo a sequência 5S do rDNA de <i>K. phaffii</i> .
pPCV-NTS	Vetor construído neste trabalho derivado do vetor pPCV (Janner <i>et al.</i> , 2013), contendo a sequência NTS do rDNA de <i>K. phaffii</i> .



pBSChr2ld-GFP	Vetor construído neste trabalho para a expressão de GFP contendo o alelo <i>leu2-d</i> como marca de seleção e uma sequência repetida do cromossomo 2 de <i>K. phaffii</i> .
pBSChr3ld-GFP	Vetor construído neste trabalho para a expressão de GFP contendo o alelo <i>leu2-d</i> como marca de seleção e uma sequência repetida do cromossomo 3 de <i>K. phaffii</i> .
pBS5Sld-GFP	Vetor construído neste trabalho para a expressão de GFP contendo o alelo <i>leu2-d</i> como marca de seleção e a sequência 5S do rDNA de <i>K. phaffii</i> .
pPCVNTSld-GFP	Vetor construído neste trabalho para a expressão de GFP contendo o alelo <i>leu2-d</i> como marca de seleção e a sequência NTS do rDNA de <i>K. phaffii</i> .
pNTSAMY-GCW	Vetor de expressão para <i>K. phaffii</i> construído em nosso laboratório para expressar amilase ancorada à parede celular. Contém a marca <i>leu2-d</i> e o promotor <i>PGK1</i> .
pBSChr2ld-AMY	Vetor construído neste trabalho para a expressão de amilase contendo o alelo <i>leu2-d</i> como marca de seleção e uma sequência repetida do cromossomo 2 de <i>K. phaffii</i> .
pBSChr3ld-AMY	Vetor construído neste trabalho para a expressão de amilase contendo o alelo <i>leu2-d</i> como marca de seleção e uma sequência repetida do cromossomo 3 de <i>K. phaffii</i> .
pBS5Sld-AMY	Vetor construído neste trabalho para a expressão de amilase contendo o alelo <i>leu2-d</i> como marca de seleção e a sequência 5S do rDNA de <i>K. phaffii</i> .
pPCVNTSld-AMY	Vetor construído neste trabalho para a expressão de amilase contendo o alelo <i>leu2-d</i> como marca de seleção e a sequência NTS do rDNA de <i>K. phaffii</i> .

## 4.2 Métodos

### 4.2.1 Análise bioinformática

A partir de dados do genoma sequenciado da linhagem de *K. phaffii* GS115 (GenBank: FN392319, FN392320, FN392321 e FN392322) foi feita uma busca por sequências repetidas, possíveis alvos de integração. Primeiramente, foi feita a busca por sequências repetidas *in tandem* usando a ferramenta Tandem Repeat Finder (Benson, 1999).

Com relação ao 5S rRNA, foi utilizada a sequência do gene homólogo de *S. cerevisiae* para identificar sequências correspondentes no genoma de *K. phaffii* usando a ferramenta Blastn.

### 4.2.2 Amplificação por PCR

Os *primers* utilizados nesse trabalho estão listados na Tabela 2. As amplificações foram realizadas utilizando-se Phusion High-Fidelity DNA Polymerase (Thermo Scientific) de acordo com as instruções do fabricante.

A tabela 5 apresenta os fragmentos amplificados neste trabalho, assim como os *primers*, o DNA molde e a temperatura de anelamento usados para cada amplicon.

**Tabela 5.** Fragmentos amplificados por PCR e condições da amplificação

<b>Amplicon</b>	<b>Par de primers*</b>	<b>Molde</b>	<b>Temperatura anelamento</b>
Alelo <i>leu2-d</i>	5-leud 3-leud	DNA genômico <i>S. cerevisiae</i>	60°C
Cassete PGK-GFP- <i>leu2d</i>	PGK- <i>chr2</i> <i>leud-chr2</i>	Vetor pK1d-GFP	55°C
Cassete PGK-GFP- <i>leu2d</i>	PGK- <i>chr3</i> <i>leud-chr3</i>	Vetor pK1d-GFP	60°C
Cassete PGK-GFP- <i>leu2d</i>	PGK-5S <i>leud-5S</i>	Vetor pK1d-GFP	60°C
Cassete PGK-GFP- <i>leu2d</i>	PGK-NTS <i>leud-NTS</i>	Vetor pK1d-GFP	58°C

NTS rDNA	18SF-pPCV 25SR-pPCV	DNA genômico <i>K. phaffii</i>	58°C
Fragmento confirmação clonagem	PGK-chr2 EGFP-Not	Vetor pBSChr2ld-GFP	58°C
Fragmento confirmação clonagem	PGK-chr3 EGFP-Not	Vetor pBSChr3ld-GFP	58°C
Fragmento confirmação clonagem	PGK-5S EGFP-Not	Vetor pBS5Sld-GFP	58°C
Fragmento confirmação clonagem	25SR-pPCV 3-leud	Vetor pPCVNTSld-GFP	58°C
Sonda EGFP	EGFP-Eco EGFP-Not	Vetor pKld-GFP	60°C
Gene <i>AMY1</i>	Falphaif-F AMYif-R	Vetor pNTSAMY-GCW	60°C

\* Ver tabela 2 (página 28)

#### 4.2.3 Preparação de células bacterianas competentes

Células de *E. coli* XL10-gold foram semeadas em meio LB ágar a partir de um estoque armazenado a -80°C e incubadas a 37°C durante a noite. Uma colônia isolada foi inoculada em 10 mL de meio SOB em um frasco Erlenmeyer de 125 mL e incubada a 37°C sob agitação a 250 rpm por 16 h. Dois mililitros desse pré-inóculo foram inoculados em 100 mL de meio SOB em um frasco Erlenmeyer de 500 mL e incubados a 37°C sob agitação a 250 rpm por aproximadamente 1 h até atingir uma OD<sub>600</sub> de 0,3. A cultura foi resfriada por 15 min em banho de água-gelo e posteriormente as células foram coletadas a 3.000 x g por 5 min a 4°C. O precipitado foi ressuscitado em 32 mL de tampão de transformação I e incubado no gelo por 15 min. Em seguida, as células foram concentradas por centrifugação sob as mesmas condições e ressuscitadas em 4 mL de tampão de transformação II. Finalmente, foram feitas alíquotas de 100 µL em tubos de 1,5 mL e armazenadas a -80°C.

#### 4.2.4 Manipulação do DNA

As digestões de DNA foram realizadas segundo as instruções do fabricante das enzimas. A tabela 3 apresenta os tampões e temperaturas recomendadas para cada enzima. Todas as reações foram incubadas por 2 horas.

Para retirar sais e enzimas após digestão o DNA foi precipitado adicionando acetato de sódio para uma concentração final de 0,3 M e 2,5 volumes de etanol 100% gelado. Após incubação a -20 °C por 16 h, o sistema foi coletado por centrifugação a 10.000 x g por 10 min. O sobrenadante foi descartado e o precipitado lavado duas vezes com 500 µL de etanol 70% por centrifugação a 10.000 x g por 2 min. O sobrenadante foi descartado e, uma vez seco, o precipitado foi ressuscitado em 10 µL de água milli Q.

A análise de fragmentos de DNA foi feita utilizando a técnica de eletroforese em gel de agarose adaptada de Sambrook *et al* (1989). O gel preparado na concentração de 1% (p/v) de agarose em tampão TAE 1X continha brometo de etídeo na concentração final de 0,5 µg/mL. As amostras foram aplicadas no gel e submetidas à corrente elétrica para permitir a migração e separação dos fragmentos de DNA. A visualização das bandas de DNA foi feita com a exposição do gel à luz ultravioleta.

Os fragmentos de DNA eluídos de gel de agarose foram purificados com o kit Wizard SV Gel and PCR Clean-Up System (Promega, EUA) seguindo as orientações especificadas pelo fabricante.

#### 4.2.5 Construção de plasmídeos

A ligação de produtos de PCR para a construção do plasmídeo contendo a marca auxotrófica *leu2-d* foi feita utilizando a enzima T4 DNA ligase (Promega) com o tampão fornecido pelo fabricante. No caso de fragmentos com extremidades coesivas o sistema foi incubado a 4°C durante 16 h e, no caso de fragmentos com extremidades abruptas, a incubação foi a 16°C pelo mesmo período de tempo.

O sistema de ligação foi utilizado para transformar células bacterianas competentes pelo método de choque térmico. Uma alíquota de células competentes armazenada a -80°C foi descongelada no gelo e em seguida foram adicionados 10 µL do sistema de ligação. As células foram incubadas no gelo por 30 min e posteriormente foram submetidas a choque térmico a 42°C por 90 s. Foram adicionados 900 µL de meio LB ao sistema que foi incubado

a 37°C por 1 h. As células foram semeadas em meio LB ágar contendo o antibiótico adequado.

A construção dos plasmídeos contendo sequências repetitivas foi feita utilizando o kit In-Fusion® HD Cloning Kit (Clontech, EUA) segundo as indicações do fabricante.

#### **4.2.6 Preparação de DNA plasmidial**

Para extrair os plasmídeos das células bacterianas em pequena escala (miniprep) foi utilizado o protocolo descrito por Sambrook *et al.* (1989) com adaptações. Cinco mililitros de meio LB contendo o antibiótico adequado foram inoculados com uma colônia transformante e incubados a 37°C por 16-18 h sob agitação. Três mililitros da cultura foram coletados por centrifugação a 10.000 x g por 2 min e o precipitado de células foi ressuspenso em 200 µL de solução I. Em seguida foram adicionados 360 µL de solução II recém-preparada, a mistura foi homogeneizada por inversão e incubada a temperatura ambiente por 5 min. Posteriormente, foram adicionados 300 µL de solução III gelada, a mistura homogeneizada e incubada no gelo por 5 min. Após esse tempo a amostra foi concentrada por centrifugação a 10.000 x g por 5 min. O sobrenadante foi transferido para outro tubo e foram adicionados 750 µL de isopropanol. A mistura foi homogeneizada por inversão e submetida a centrifugação a 10.000 x g por 5 min. O sobrenadante foi descartado e o precipitado ressuspenso em 200 µL de solução I. Em seguida, foram adicionados 110 µL de acetato de amônio 7,5 M e a mistura homogeneizada vigorosamente seguido de centrifugação a 10.000 x g por 10 min. O sobrenadante foi transferido para outro tubo e foram adicionados 750 µL de etanol 100% gelado. Após centrifugação a 10.000 x g por 5 min, o sobrenadante foi descartado e o precipitado lavado com 500 µL de etanol 70% seguido de uma nova centrifugação a 10.000 x g por 2 min. O sobrenadante foi descartado e o excesso de etanol foi seco em um concentrador a vácuo. O precipitado foi ressuspenso em 50 µL de água milli Q contendo RNase A (0,1 mg/mL) e incubado a 37°C por 15 min. O DNA foi armazenado a -20°C até seu uso.

Para a obtenção dos plasmídeos em maior quantidade foi utilizado o kit Qiagen Plasmid Maxi (Qiagen) seguindo as orientações do fabricante.

#### **4.2.7 Transformação de leveduras por eletroporação**

A transformação com vetores integrativos foi feita por eletroporação segundo o protocolo descrito no EasySelect Pichia Expression Kit (Invitrogen, EUA) com algumas modificações. Uma colônia fresca de levedura foi inoculada em 5 mL de YPD. Após 12 h de crescimento a 28°C sob agitação de 250 rpm, 500 µL desse pré-inóculo foram usados para inocular mais 5 mL de meio. Após 8 h de incubação, o volume necessário dessa nova cultura foi inoculado em 100 mL de meio YPD para atingir uma OD<sub>600</sub> de 1,3 após 16 h de crescimento. As células foram coletadas e lavadas 3 vezes com água estéril gelada (100 mL na primeira lavagem e 50 mL nas outras duas) e uma vez com sorbitol 1 M gelado centrifugando-se a 1.500 x g por 5 min a 4°C. Posteriormente, as células foram ressuspensas em 300 µL de sorbitol 1 M gelado e 80 µL dessas células competentes foram misturadas com 5-10 µg de DNA linearizado e transferidas a uma cubeta de eletroporação de 2 mm (Bio-Rad). A cubeta foi incubada no gelo por 5 min antes de aplicar um pulso elétrico sob as seguintes condições: 1500 V, 400 Ω, e 25 µF. Imediatamente após o pulso, 1 mL de sorbitol 1 M gelado foi adicionado. Quando a marca de seleção foi dominante, as células foram incubadas a 30°C sem agitação durante 2 h antes de serem semeadas no meio contendo o antibiótico adequado. No caso de marcas auxotróficas as células foram semeadas diretamente em meio mínimo para seleção de transformantes.

#### **4.2.8 Extração de DNA total de leveduras**

O DNA genômico de *K. phaffii* foi extraído com o protocolo descrito a seguir. As células foram crescidas em 40 mL de meio MD em frascos de 250 mL a 30°C sob agitação de 200 rpm durante 24 h. Depois as células foram coletadas por centrifugação a 3.000 x g por 5 min e ressuspensas em 3 mL de tampão SE. Posteriormente 100 µL de uma solução de liticase 20 mg/mL foram adicionados e a mistura foi incubada a 37°C por 1 h seguido de centrifugação a 3.000 x g por 5 min. As células foram ressuspensas em 3 mL de TE<sub>20</sub> e foram adicionados 500 µL de SDS 10% e a mistura foi incubada a 65°C por 30 min. Posteriormente foram adicionados 1,5 mL de acetato de potássio 5 M, pH 8,9 e a mistura foi incubada no gelo por 30 min seguido de centrifugação a 10.000 x g por 10 min a 4°C. O sobrenadante foi transferido para um tubo novo e foi adicionado 1 volume de clorofane seguido de centrifugação a 3.000 x g por 10 min. A fase aquosa foi transferida para um tubo novo e foram adicionados 2,5 volumes de etanol 100%. Após 5 min de incubação a

temperatura ambiente a mistura foi sedimentada a 1.500 x g por 10 min. O precipitado foi lavado com 500 µL de etanol 70% e submetido a centrifugação a 1.500 x g por 2 min. Finalmente o precipitado foi ressuspensionado em 100 µL de TE contendo RNase A (0,1 mg/mL).

#### 4.2.9 Southern blot

Aproximadamente 10 µg de DNA genômico foram digeridos com BglII, EcoRI ou EcoRV a 37°C por 16 h e o DNA digerido foi aplicado em gel de agarose 1%.

O gel foi tratado com a solução de depurinação por 15 min, seguindo-se duas lavagens com a solução de desnaturação por 15 min e, finalmente, duas lavagens com a solução de neutralização por 30 min. Depois o gel foi transferido para a membrana de nitrocelulose como descrito por Sambrook *et al.*, 1989.

A marcação da sonda, a hibridização e a revelação foram feitas com o kit AlkPhos Direct Labeling and CDP-Star Detection System (GE Life Sciences) conforme as especificações do fabricante. As sondas utilizadas foram:

- Um fragmento de ~600 pb correspondente ao promotor *PGK1* (Arruda *et al.*, 2016) obtido por digestão do vetor pKId-GFP com as enzimas BglII e BamHI (Anexo 1).
- Um fragmento de ~750 pb correspondente ao gene EGFP obtido por PCR com os *primers* EGFP-Eco e EGFP-Not (Anexo 2).

A temperatura de hibridização foi de 55°C. A detecção por quimioluminescência foi feita no equipamento Amersham Imager 600 (GE Life Sciences) e a intensidade das bandas foi medida com o software Image-Quant TL 8.1.

#### 4.2.10 Eletroforese em campo pulsado

Os cromossomos de *K. phaffii* foram separados por eletroforese em campo pulsado, como descrito por Ohi *et al.* 1998 com algumas modificações. As células foram crescidas em 20 mL meio MD até a saturação. Dois mililitros da cultura foram concentrados por centrifugação a 1.200 x g por 5 min para ter ~100 µL de precipitado e este foi lavado com 1 mL de EDTA 0,05 M pH 8,0 e submetido a centrifugação a 1.200 x g por 5 min. O precipitado foi ressuspensionado em 300 µL de EDTA 0,05 M pH 8,0 contendo DTT 0,05 M e foi incubado a 30°C por 20 min. A mistura foi submetida a centrifugação a 1.200 x g por 5

min e o precipitado foi ressuspensionado em 490 mL de EDTA 0,05 M pH 8,0. Posteriormente, foram adicionados 10 µL de liticase e 500 µL de agarose 1% (Agarose for Pulsed Field Electrophoresis: Sample Preparation, Sigma) e a mistura foi colocada nos moldes fornecidos pelo sistema CHEF Genomic DNA Plug kit (BioRad) até solidificar a 4°C. Os blocos de agarose foram incubados em solução de esferoplastos a 37°C durante a noite. Seguidamente, foram lavados com EDTA 0,05 M pH 8,0 e incubados em solução de protease a 50°C por três dias. Finalmente, foram lavados com EDTA 0,05 M pH 8,0 e armazenados nessa solução a 4°C até o uso.

O equipamento usado para realizar a eletroforese em campo pulsado foi o CHEF DRIII System (BioRad). O tampão de corrida foi tampão TBE 0,5X trocado a cada 24 h. A separação foi feita em gel de agarose 1% (Agarose for Pulsed Field Electrophoresis: Running Gel, Sigma) preparado no mesmo tampão. A eletroforese foi realizada a 14°C com uma voltagem de 2,5 V/cm e um ângulo de 106°. Os intervalos de alternância foram: 300 s por 24 h, 600 s por 24 h e 900 s por 54 h. Depois da separação o gel foi incubado em solução de brometo de etídeo para visualizar o DNA na luz ultravioleta. Posteriormente o gel foi tratado com as soluções de depurinação, desnaturação e neutralização e o DNA foi transferido a uma membrana de nitrocelulose para análise por Southern blot.

#### **4.2.11 Análise de fluorescência no Typhoon scanner**

As células foram crescidas em meio sólido (MD ágar) durante 2-3 dias. A placa foi analisada utilizando o Typhoon Scanner (GE) sob as seguintes condições: Filtro de emissão: 526 fluoresceína, Cy2, AlexaFluor 488; PMT: 600; Laser: verde (532); Sensibilidade: normal; Tamanho do pixel: 200 µm; Pannel focal: +3 mm.

#### **4.2.12 Citometria de fluxo**

As células foram crescidas em 5 mL de meio MD durante 24 h e posteriormente o volume adequado de cada pré-inóculo foi adicionado em mais 5 mL de meio MD para obter uma OD<sub>600</sub> de 0,5 após 24 h de crescimento. Posteriormente as células foram lavadas 2 vezes com PBS contendo Tween 0,5% centrifugando a 3000 x g por 5 min a 4°C. Finalmente, as células foram ressuspensionadas no volume adequado de PBS a fim de ter aproximadamente



$1 \times 10^6$  células/mL e foram mantidas a 4°C até a análise no citômetro FACSVerse. A análise dos dados adquiridos foi feita com o software FlowJo.

#### **4.2.13 Eletroforese em gel de poliacrilamida**

As proteínas foram analisadas por eletroforese em gel de poliacrilamida. No caso do gel desnaturante (SDS-PAGE) as amostras foram fervidas por 5 min antes da aplicação no gel. A corrida foi conduzida em tampão de corrida 1X com uma voltagem de 100 V. No caso do gel não desnaturante, a corrida foi realizada a 4°C.

As bandas de proteínas foram visualizadas após a coloração com Coomassie blue. O gel foi incubado na solução corante por 16 h sob agitação, seguido de incubação na solução descorante, sob agitação, até o aparecimento das bandas.

#### **4.2.14 Análise de fluorescência em gel de poliacrilamida não desnaturante**

As células foram crescidas em 5 mL de meio MD durante 24 h e posteriormente o volume adequado de cada pré-inóculo foi adicionado em 20 mL de meio MD para começar a cultura com uma OD<sub>600</sub> de 0,03. Após 20 h de incubação a 30°C e 200 rpm, 2 mL da cultura foram concentrados por centrifugação a 13.000 x g durante 3 min e as células foram lavadas com PBS. O precipitado foi pesado (peso úmido) e as células foram ressuspendidas no volume necessário de tampão Tris-HCl 50 mM pH 8,0 (4,5 mL de tampão para cada 1 g de peso úmido de células) contendo Tween 20 na concentração de 1,25% e 30 µL de mix de inibidores de proteases por cada 1 mL de tampão (solução estoque do mix de inibidores: 1 tablete de Complete Tablets Mini EDTA-free Easy pack dissolvida em 1 mL de água destilada). Posteriormente, foi adicionada a quantidade necessária de esferas de vidro de 0,5 mm de diâmetro (4 g para cada 1 g de peso úmido de células). A mistura foi submetida a agitação vigorosa (vortex) por 30 s, seguido de incubação no gelo por 30 s, e este procedimento foi repetido 30 vezes. A parte líquida foi transferida para um tubo e concentrada por centrifugação a 10.000 x g por 30 min a 4°C. O sobrenadante foi transferido para um tubo e foi armazenado a -20°C até o uso.

Doze microlitros de sobrenadante foram aplicados em gel de poliacrilamida (SDS-PAGE). Com base na intensidade das bandas de proteínas observadas no gel foi determinada a quantidade de sobrenadante a ser aplicada no gel não desnaturante para garantir a mesma

quantidade de extrato proteico para todos os clones. A fluorescência da GFP foi detectada no equipamento Amersham Imager 600. A fluorescência de cada clone foi comparada com a de um clone controle contendo uma única cópia do gene EGFP para obter um valor de produção de GFP relativo.

#### **4.2.15 Detecção da atividade amilase em placa**

As células transformantes foram crescidas em meio mínimo com amido 1% durante 48 h a 30°C. Depois a placa foi corada com vapor de iodo até o aparecimento de halos transparentes devido à hidrólise do amido presente no meio.

#### **4.2.16 Detecção da atividade amilase em meio líquido**

As células transformantes foram crescidas em 10 mL de meio YPD ou meio MD a 28°C e 200 rpm durante 24 h e posteriormente o volume adequado de cada pré-inóculo foi adicionado em 50 mL do respectivo meio para começar a cultura com uma OD<sub>600</sub> de 0,1 no caso do meio YPD e de 0,02 no caso do meio MD. Antes de transferir o pré-inóculo para o meio novo, a cultura foi lavada com água destilada estéril para retirar a amilase que estivesse presente no meio. Os inóculos foram incubados a 28°C e 200 rpm durante 68 h no caso do meio YPD e foram tirados 2 mL de amostra às 14 h, 22 h, 40 h e 68 h. No caso do meio MD os inóculos foram incubados durante 72 h e a cada 24 h foram tirados 2 mL de amostra. Durante o ensaio o crescimento celular foi acompanhado pela dosagem da OD<sub>600</sub>.

As amostras coletadas do cultivo foram concentradas por centrifugação a 3.000 x g durante 5 min e o sobrenadante foi empregado no ensaio enzimático de degradação de amido.

#### **4.2.17 Determinação de biomassa**

*K. phaffii* M12 foi inoculada em um frasco de 1 L contendo 100 mL de meio YPD e incubada a 30°C e 200 rpm durante 12 h. A cada 2 horas foi dosada a OD<sub>600</sub> da cultura e foram coletados 4,5 mL de amostra e divididos por igual em três tubos Eppendorf. As amostras foram concentradas por centrifugação a 3.000 x g durante 5 min e o precipitado foi seco a 60°C até não ter mais variação do peso. Finalmente, foi gerada uma curva de calibração relacionando a OD<sub>600</sub> com o peso seco da levedura.

#### **4.2.18 Ensaio enzimático de degradação de amido**

A atividade amilolítica foi determinada pelo método descrito por Fuwa (1954) com algumas modificações. Sessenta microlitros de sobrenadante foram misturados com 40  $\mu$ L de tampão acetato de sódio 0,5 M pH 6,0 e com 100  $\mu$ L de amido 0,5% (p/v). A reação foi incubada a 40°C por 30 min. No caso das amostras com maior concentração de enzima foi diminuído o volume de sobrenadante usado e/ou o tempo de reação. Após incubação, foram adicionados 200  $\mu$ L de ácido acético 1 M para parar a reação. Duzentos microlitros do reagente de FUWA e 4,4 mL de água destilada foram adicionados, a mistura foi homogeneizada e 200  $\mu$ L foram usados para ler a absorbância a 660 nm. O valor obtido foi comparado com uma curva de calibração feita com amido solúvel na faixa 0,1-0,5 mg. Foram feitas triplicatas tanto das amostras como da curva de calibração.

Uma unidade de atividade amilolítica foi definida como a quantidade de enzima que hidrolisa 0,1 mg de amido por minuto. A atividade foi expressada em termos de U/mL levando em consideração o volume de amostra (sobrenadante) usado na reação. Finalmente este último valor foi dividido pelo valor da biomassa (mg/mL) medido a cada tempo do ensaio, obtendo-se um valor de atividade específica relacionado com a concentração celular (U/mg), o qual é comparável entre os diferentes clones.

#### **4.2.19 Detecção de atividade amilolítica em gel de poliacrilamida – zimograma**

As células foram crescidas em 10 mL de meio mínimo com glicerol 1% como fonte de carbono durante 24 h e posteriormente o volume adequado de cada pré-inóculo foi adicionado em 100 mL do mesmo meio para começar a cultura com uma OD<sub>600</sub> de 0,01. Após 48 h de incubação a 30°C e 200 rpm, 40 mL da cultura foram submetidos a centrifugação a 13.000 x g durante 3 min e foram recuperados tanto o sobrenadante quanto o precipitado. As células foram lavadas com 2 mL de PBS, submetidas ao protocolo de lise celular descrito no item 4.2.14 e finalmente o precipitado foi ressuscitado em tampão de amostra não desnaturante. O sobrenadante da cultura foi precipitado com TCA como descrito a continuação. Para cada 1 mL do sobrenadante foram adicionados 250  $\mu$ L de TCA 100%. Após incubação a 4°C durante 16 h, a mistura foi submetida a centrifugação a 8.000 x g por 30 min a 4°C. O precipitado foi lavado com 1 mL de acetona gelada e novamente submetido a centrifugação a 8.000 x g por 15 min. A lavagem com acetona foi repetida e depois de deixar secar, o precipitado foi ressuscitado em tampão de amostra não desnaturante e armazenado a -20 °C até análise em

gel SDS-PAGE. Para determinação da atividade amilolítica foram usadas ambas as amostras: extrato proteico intracelular e o sobrenadante da cultura. As amostras foram aplicadas em gel de poliacrilamida não desnaturante (10%) nas condições descritas no item 4.2.13. Após a corrida, o gel foi incubado com tampão acetato de sódio 50 mM pH 5,5 por 1 h. Posteriormente o tampão foi trocado por uma solução de amido 0,5% dissolvida em tampão acetato de sódio 50 mM pH 5,5 e o gel foi incubado a 4°C por 12 h. Após esse período, o gel foi incubado a 37°C por 2 h. Finalmente essa solução foi trocada pela solução reveladora para visualização das bandas de degradação de amido.

#### 4.2.20 Teste de estabilidade genética

A estabilidade do DNA integrado no genoma da levedura foi testada após 72 gerações. Uma colônia foi crescida em 10 mL de meio YPD a 30°C e 200 rpm. Depois de 24 h, 400 µL do pré-inóculo foram inoculados em 40 mL de YPD e incubados sob as mesmas condições durante 24 h. Posteriormente, 400 µL da cultura foram transferidos para outro frasco com 40 mL de YPD e incubados sob as mesmas condições durante 24 h. Este procedimento foi repetido quatro vezes mais para um tempo total de crescimento de 144 h. Após esse tempo (72 gerações), o DNA genômico foi extraído e analisado por Southern blot.

#### 4.2.21 Avaliação do crescimento em meio líquido

Uma colônia nova da levedura foi inoculada em 500 µL de MD em placa *deep-well* e incubada a 30°C e 250 rpm por 24 h. Um volume de 100 µL de MD foi inoculado com o volume necessário do pré-inóculo para começar a cultura com uma OD<sub>600</sub> de 0,1 em placa de 96 poços. A análise de crescimento foi realizada com o espectrofotômetro Eon High Performance Microplate Spectrophotometer (Biotek) incubando a 30°C sob agitação de 300 rpm por 72 h, coletando dados de OD<sub>600</sub> a cada 30 minutos.

Durante a fase exponencial de crescimento a velocidade específica de crescimento ( $\mu$ ) é constante e máxima. Dessa forma, a velocidade máxima específica de crescimento  $\mu_{\text{máx}}$  pode se expressar como:

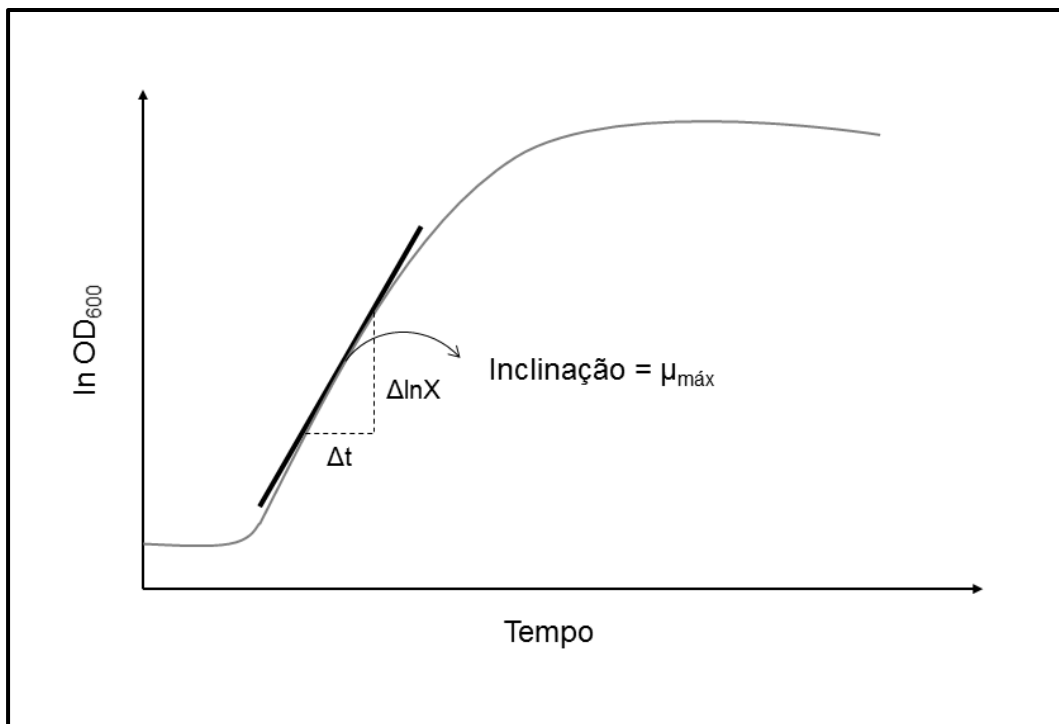
$$\mu_{\text{máx}} = \frac{dX}{dt} \frac{1}{X}$$

Após integrar esta equação é possível determinar  $\mu_{\text{máx}}$  como:

$$\mu_{\text{máx}} = \frac{\ln X_f - \ln X_i}{t_f - t_i}$$

Onde  $X_f$  é a concentração celular no final da fase exponencial,  $X_i$  é a concentração celular no início da fase exponencial,  $t_f$  é o tempo final da fase exponencial e  $t_i$  é o tempo inicial da fase exponencial.

Para plotar as curvas de crescimento foram usados os valores de logaritmo natural da  $OD_{600}$  e a velocidade máxima de crescimento foi calculada como a inclinação da parte linear dessas curvas que corresponde à fase exponencial de crescimento (figura 5).



**Figura 5.** Cálculo da velocidade máxima específica de crescimento ( $\mu_{\text{máx}}$ ). A parte linear na curva de crescimento corresponde à fase exponencial e a inclinação dessa linha corresponde a  $\mu_{\text{máx}}$ .  $\Delta \ln X$ : variação do logaritmo natural da concentração celular entre o ponto inicial e o ponto final da fase exponencial,  $\Delta t$ : variação de tempo entre o ponto inicial e o ponto final da fase exponencial.

#### 4.2.22 Análise de dados

As análises estatísticas e os gráficos foram feitos com o software GraphPad Prims 5. Foi aplicado ANOVA seguido do teste de Tukey para comparação entre as amostras. As barras de erro nos gráficos representam o erro padrão da média.

## 5. RESULTADOS E DISCUSSÃO

### 5.1 Identificação de sequências repetidas no genoma de *K. phaffii*

A primeira etapa desse projeto foi a identificação de sequências repetidas *in tandem* no genoma de *K. phaffii* para serem possíveis alvos de integração por recombinação homóloga. Utilizando ferramentas de bioinformáticas foram identificadas 198 sequências no cromossomo 1, 175 no cromossomo 2, 144 no cromossomo 3 e 127 no cromossomo 4. Dez dessas sequências foram selecionadas e submetidas a análise com a ferramenta Blastn para identificar quais delas não fazem parte de uma ORF e, dessa forma, serem consideradas candidatas a sítio de integração de vetores. A tabela 6 apresenta as sequências selecionadas, seu tamanho e o número de vezes em que a sequência está presente no genoma. Como pode ser observado na tabela, a segunda sequência do cromossomo 2 e a primeira do cromossomo 3 (identificadas em vermelho) não codificam para nenhuma proteína e, portanto, foram selecionadas como possíveis alvos de integração.

**Tabela 6.** Sequências repetidas *in tandem* identificadas no genoma de *K. phaffii* \*

Sequência	Tamanho da repetição	Número de correspondências
<b>Cromossomo 1</b>		
GCAGAACTGGATACTTCGGAAGAAGATGGGGCA		
<i>Pichia pastoris</i> GS115 hypothetical protein (PAS_chr1-1_0135) mRNA, complete cds	33 pb	3
GTGGATTCCTCAGTAGATTCCTCAACCTCCTCAGTGGA AGTAGACTCCTCAACCTCATCAGTAGAT	66 pb	4
<i>Pichia pastoris</i> GS115 Mucin-like protein (PAS_FragB_0067) mRNA, complete cds		
GATTCCTCAGCCTCCTCGGTGGAGGTGGATTCCTCAGC ATCCTCAGTAGAGGTA	54 pb	8
<i>Pichia pastoris</i> GS115 Mucin-like protein (PAS_FragB_0067) mRNA, complete cds		

---

## Cromossomo 2

---

Sequência	Tamanho da repetição	Número de correspondências
CCCACCAAGCGAGCGCAGCGACGCTGGTTGGTGGACG CAAATATCCACCAGTCAACG		
<i>Pichia pastoris</i> GS115 hypothetical protein (PAS_chr2-1_0002) mRNA, complete cds	57 pb	3
GAGCCAACTGAGGAACCAACCTCTGAGTCTACTGAGG AGCCTACTGAAGAACCAACTTCTGAAGCAACTTCT		
<i>Pichia pastoris</i> GS115 chromosome 2, complete sequence	72 pb	4
<hr/>		
<h2>Cromossomo 3</h2> <hr/>		
ACCCTCGCCGCATGATCTCCCGCGGTTGGACTTGCCAT CATCTTGAGTCCCCCAGCCTCTCCGCCAGCTGCGCT CCGCTTGCCGGCC		
<i>Pichia pastoris</i> GS115 chromosome 3, complete sequence	89 pb	4
GAGTCGATAGTAGACGATGGTACCACAGCAATAGTAC TTCCATCCTCATCGGTAGTGAACGTGGTGTAAAGTACCA GTCCAGGTGGATGTA		
<i>Pichia pastoris</i> GS115 hypothetical protein (PAS_chr3_0002) mRNA, complete cds	90 pb	3
TTGTCATCTTTCTTCTCATCAGTTTTGGCACCGAATGAA AATGCTGGC		
<i>Pichia pastoris</i> GS115 Essential component of the nuclear pore complex, which mediates nuclear import and export (PAS_chr3_0825) mRNA, complete cds	48 pb	7

---

---

## Cromossomo 4

---

Sequência	Tamanho da repetição	Número de correspondências
TGACGAATGTACAGAACCTCCAGAACCTCCAGTGTGC TCTGATGACGA	48 pb	4
<i>Pichia pastoris</i> GS115 hypothetical protein (PAS_chr4_0357) mRNA, complete cds		
TGAACTGGTGGTGGCTGAACCTGCTGNTGC	30 pb	4
<i>Pichia pastoris</i> GS115 chromosome 4, complete sequence		

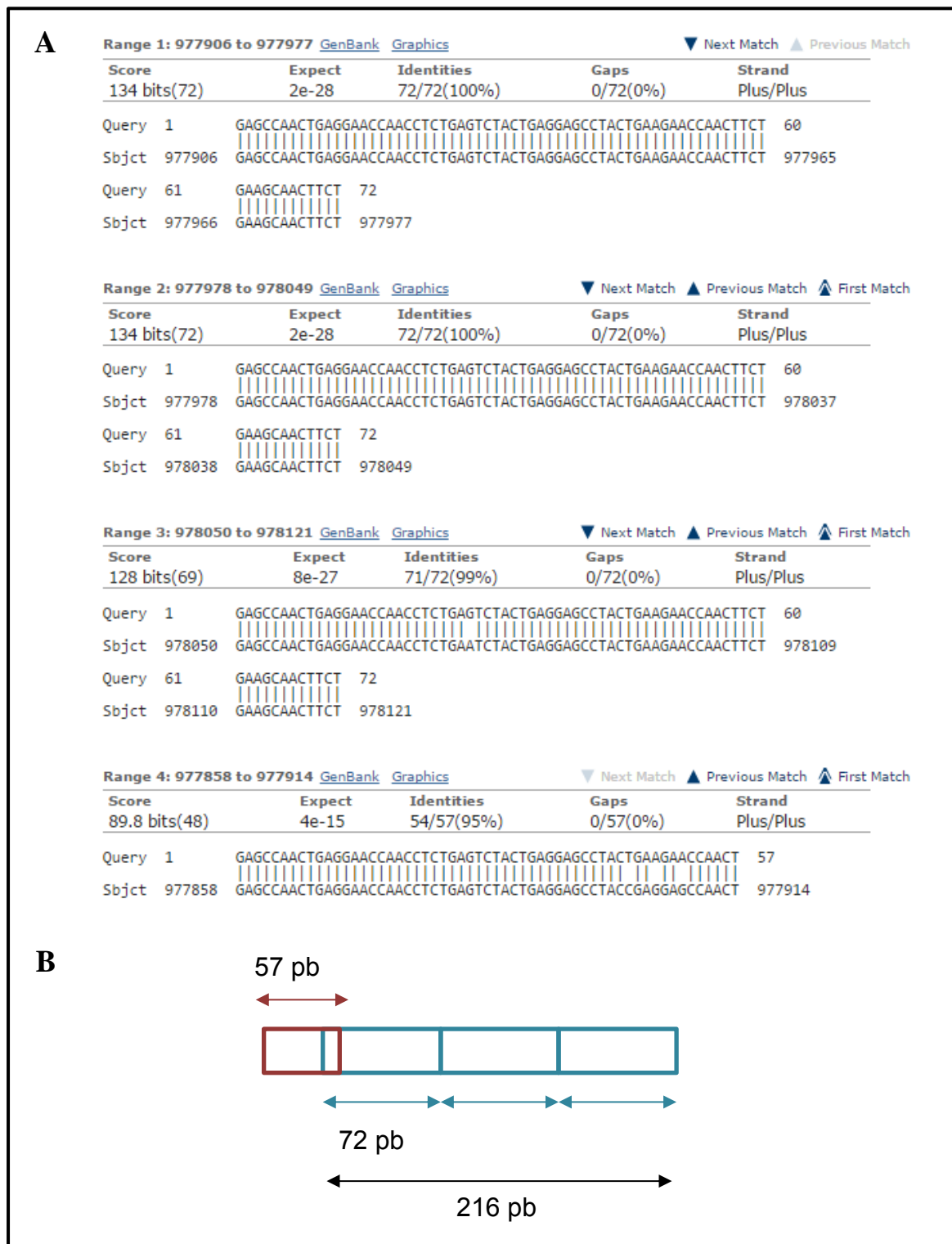
---

\*Em vermelho, sequências selecionadas como alvos de integração.

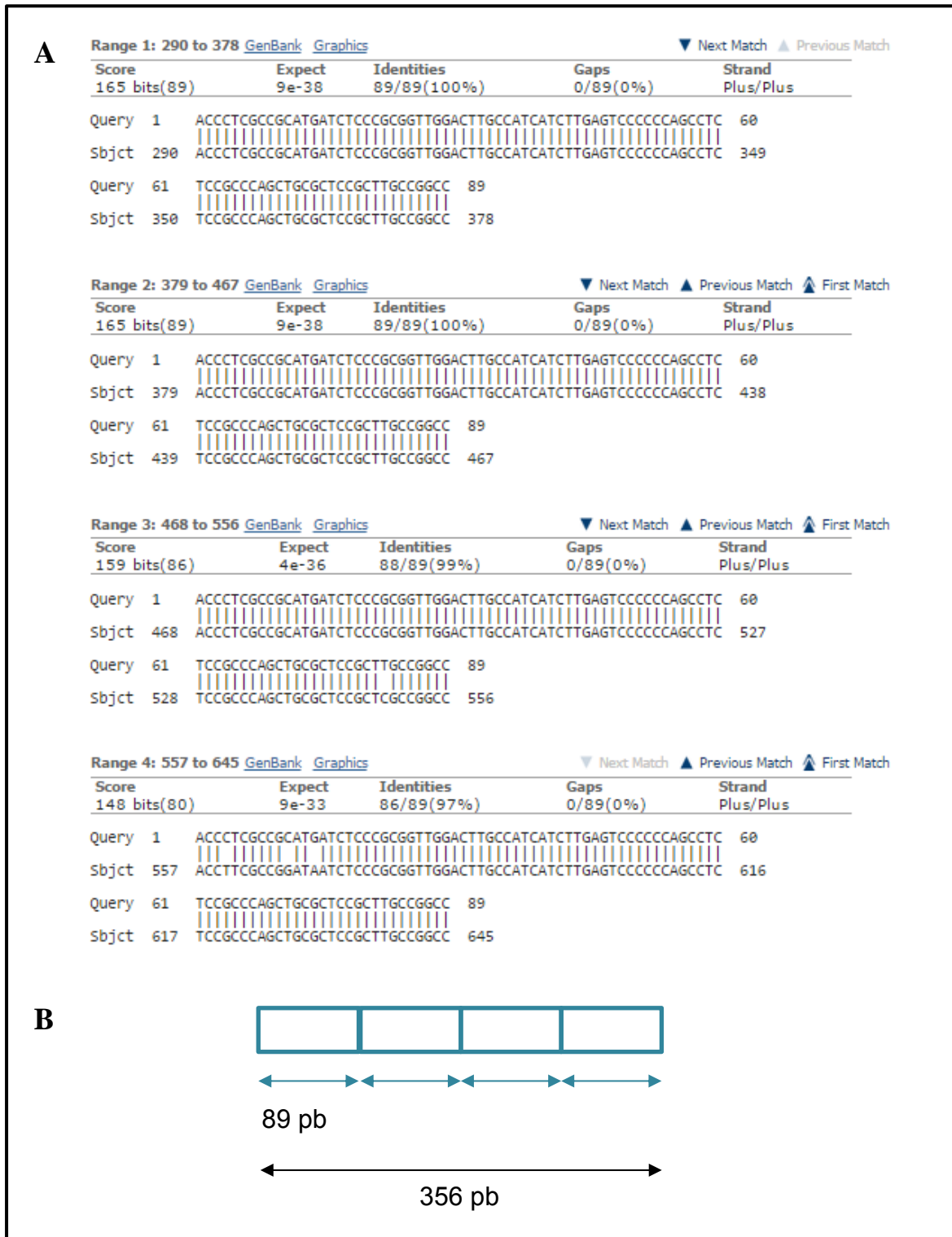
As figuras 6A e 7A apresentam o resultado obtido ao analisar com a ferramenta Blastn as sequências selecionadas do cromossomo 2 e 3, respectivamente. Com relação à sequência do cromossomo 2, das quatro correspondências encontradas, duas apresentam 100% de identidade entre si e a terceira apresenta 99% de identidade com a mudança de apenas um par de bases. Essas três repetições encontram-se *in tandem*. A quarta correspondência é uma sequência de 57 pb com 95% de identidade e que se sobrepõe com os primeiros 9 pb do *tandem*. A figura 6B mostra um esquema da organização dessas repetições no cromossomo 2.

Com relação à sequência do cromossomo 3, das quatro correspondências encontradas, também duas apresentam 100% de identidade entre si, a terceira, 99% com a mudança de apenas um par de bases e a quarta, 97% com mudança em três pares de bases. As quatro sequências encontram-se *in tandem* como esquematizado na figura 7B.



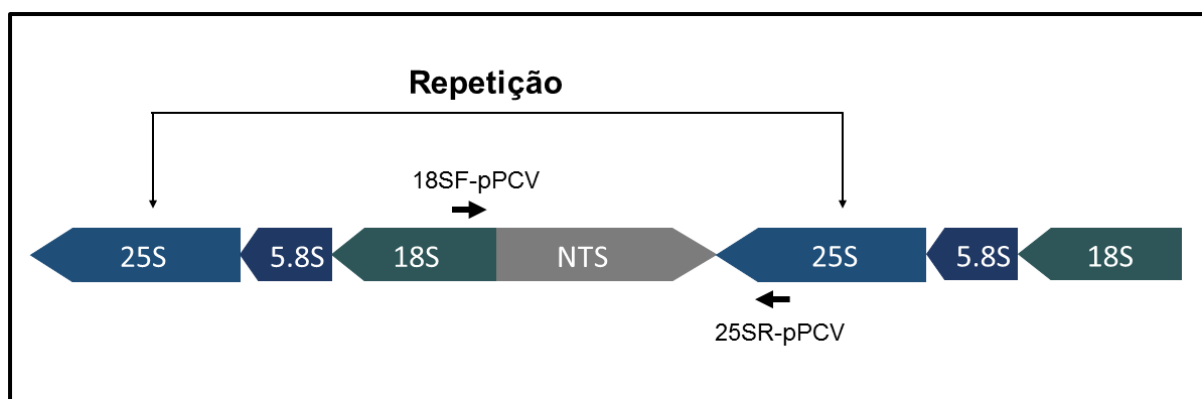


**Figura 6. Sequência alvo de integração no cromossomo 2 de *K. phaffii*. A.** Resultado da análise com a ferramenta Blastn da sequência de 72 pb previamente identificada com a ferramenta Tandem Repeat Finder, **B.** Esquema da organização das repetições.



**Figura 7. Sequência alvo de integração no cromossomo 3 de *K. phaffii*. A.** Resultado da análise com a ferramenta Blastn da sequência de 89 pb previamente identificada com a ferramenta Tandem Repeat Finder, **B.** Esquema da organização das repetições.

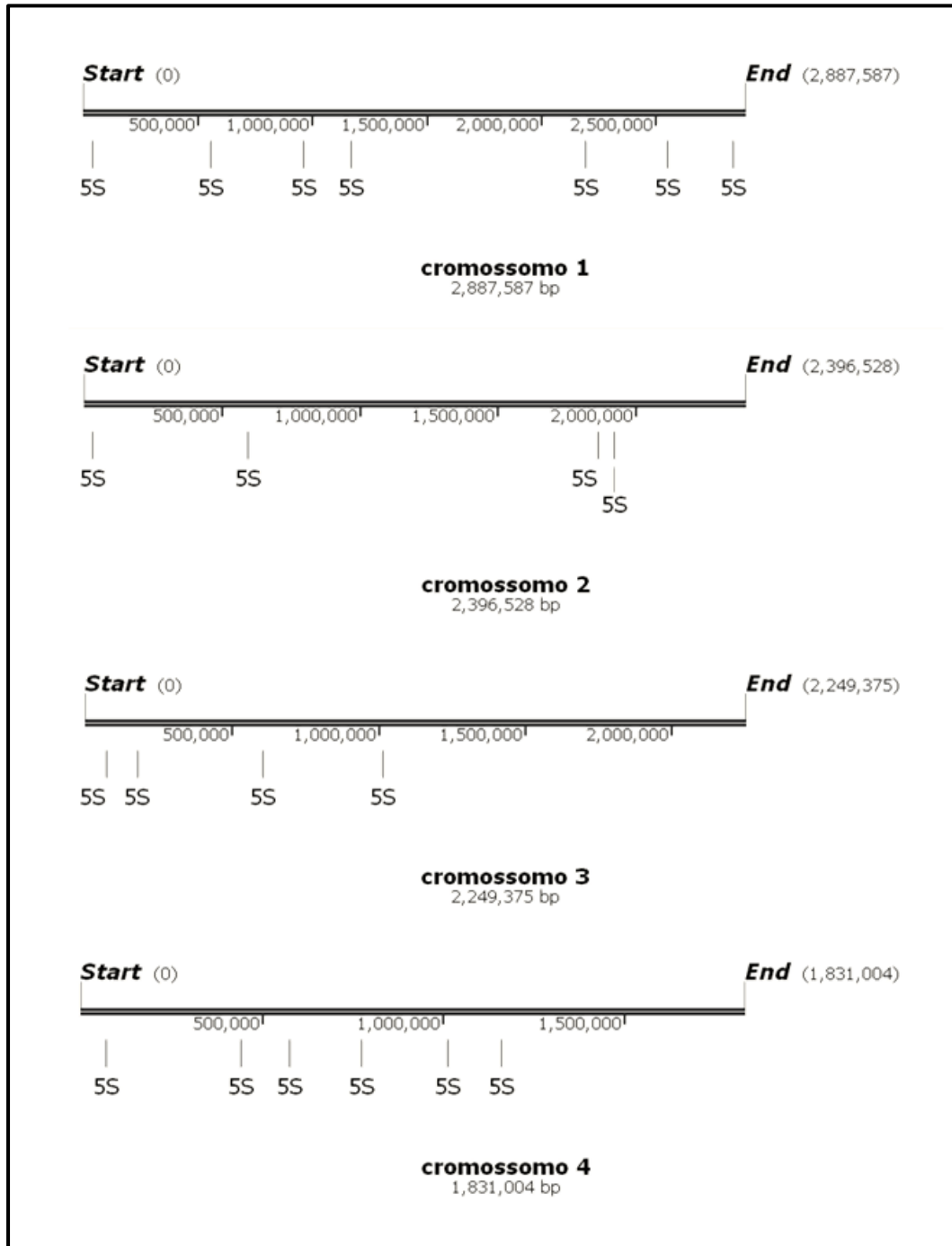
Outra sequência analisada nesse trabalho foi o espaçador não-transcrito (NTS) presente no *locus* do DNA ribossomal. A sequência NTS se repete várias vezes nesse *locus* separando as unidades 25S, 5.8S e 18S que são codificadas *in tandem*. Esta região espaçadora, de aproximadamente 2,5 kb, já foi utilizada para gerar linhagens de *K. phaffii* contendo um alto número de cópias (Marx *et al.*, 2009) e, portanto, foi escolhida como controle para comparação com as sequências aqui testadas. No presente trabalho foram utilizados os *primers* 18SF-pPCV e 25SR-pPCV (tabela 2 e figura 8) para obter um fragmento contendo ~250 pb da unidade 18S, toda a região NTS e ~290 pb da unidade 25S.



**Figura 8.** Sequência alvo de integração na região NTS do rDNA de *K. phaffii*. As setas representam os sítios onde anelam os *primers* 18SF-pPCV e 25SR-pPCV utilizados para amplificar a sequência alvo.

Com relação ao rDNA 5S, foi identificada uma sequência repetida de 121 pb que se encontra espalhada no genoma da levedura. Em *K. phaffii* GS115 haviam sido identificadas 21 cópias do rDNA 5S (De Schutter *et al.*, 2009). Uma análise usando a ferramenta Blastn mostrou que esta sequência se repete seis vezes no cromossomo 1, sendo que uma das correspondências possui 118 pb. No cromossomo 2, a sequência está repetida quatro vezes, sendo que uma das correspondências tem 99% de identidade devido à mudança de um nucleotídeo. No cromossomo 3, existem quatro repetições da sequência, sendo que uma delas tem 99% de identidade e outra possui apenas 32 pb. No cromossomo 4, foram identificadas seis repetições da sequência, sendo que uma delas possui 119 pb. Uma última correspondência foi identificada em um *contig* separado de 80073 pb. Com o novo sequenciamento do genoma da linhagem GS115 (Love *et al.*, 2016) foi possível identificar

que esta última repetição encontra-se no cromossomo 1. A figura 9 apresenta um esquema da localização da sequência rDNA 5S em cada um dos cromossomos.



**Figura 9.** Localização da sequência do rDNA 5S nos cromossomos de *K. phaffii*. A sequência de 121 pb encontra-se espalhada no genoma em 21 *loci* diferentes.

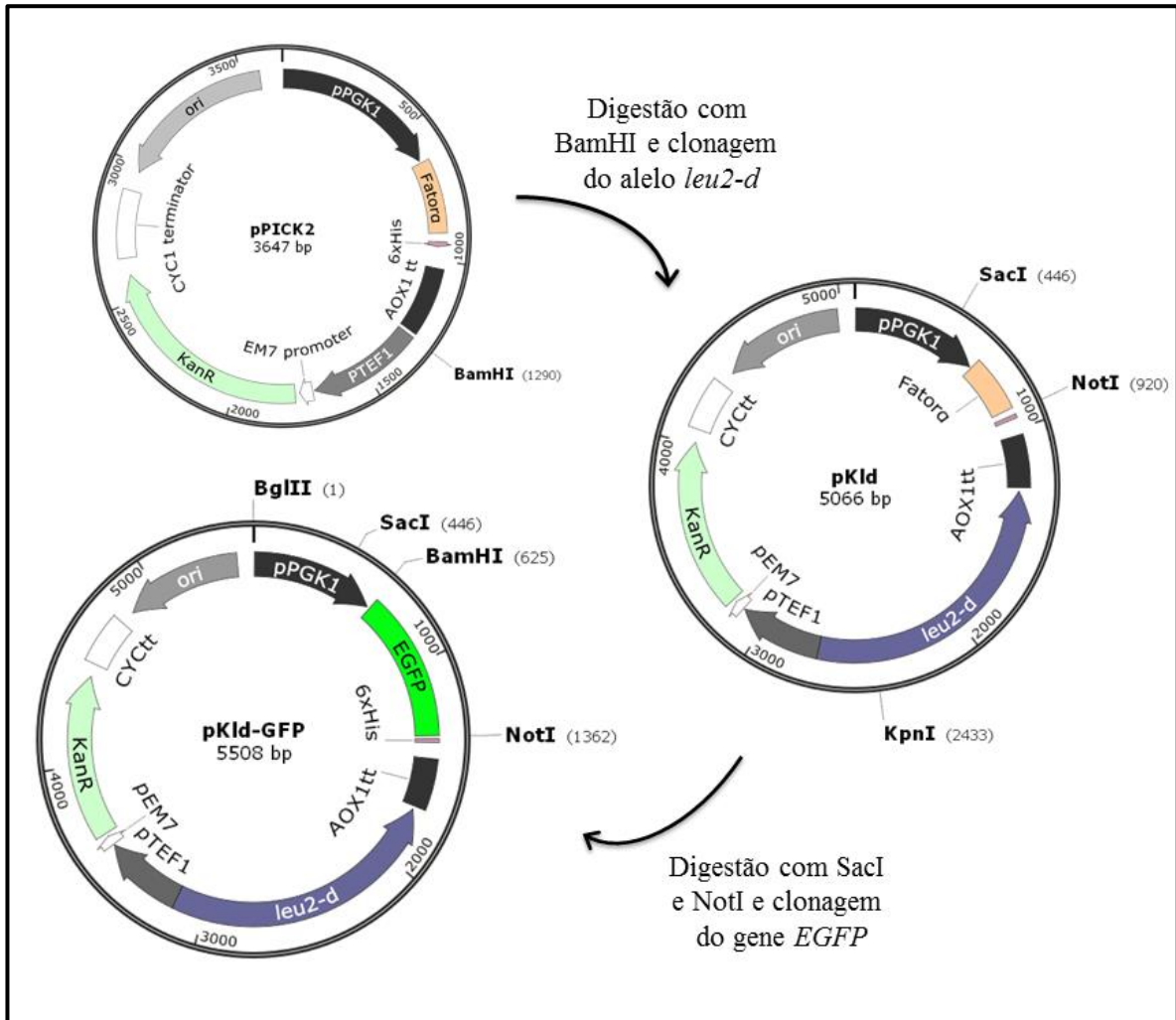
## 5.2 Integração múltipla do gene repórter EGFP

A proteína verde fluorescente (GFP: *Green Fluorescent Protein*) é amplamente usada como repórter intracelular já que tem a propriedade de fluorescer sem a adição de um indutor. GFP autocatalisa a formação do seu próprio cromóforo, o que é uma vantagem na hora de usá-la como repórter em sistemas heterólogos. A expressão de GFP em *Escherichia coli* e *Caenorhabditis elegans* mostrou que o cromóforo pode ser formado na ausência de outros produtos de *A. victoria* (Chalfie *et al.*, 1994).

Neste trabalho foi usada uma variante do gene GFP de *A. victoria* que codifica a proteína GFPmut1 que possui uma substituição dos aminoácidos Phe-64 por Leu e Ser-65 por Thr (Cormack, Valdivia e Falkow 1996). Esta dupla mutação deslocou o pico máximo de absorvância a 395 nm para 488 nm, o que permite o uso de conjuntos de filtros padrões de excitação-emissão para isotiocianato de fluoresceína (FITC) e permite a excitação com laser de argônio para a detecção de GFP por citometria de fluxo. Além disso, esta variante contém mais de 190 mutações silenciosas na região codante que correspondem a códons preferenciais em humanos. Esta versão da proteína é chamada de EGFP (*Enhanced Green Fluorescent Protein*) já que apresenta uma fluorescência mais intensa e uma maior expressão em células de mamíferos.

### 5.2.1 Construção do vetor pKId-GFP

A estratégia de construção do plasmídeo contendo a marca *leu2-d* e o gene repórter EGFP sob controle do promotor  $P_{PGK1}$  está representada na figura 10. Neste trabalho foi usado um fragmento de 618 pb do promotor  $P_{PGK1}$  de *K. phaffii* o qual é capaz de controlar a expressão de proteínas heterólogas e de promover a integração do vetor no *locus PGK1* no genoma da levedura (Arruda *et al.*, 2016).



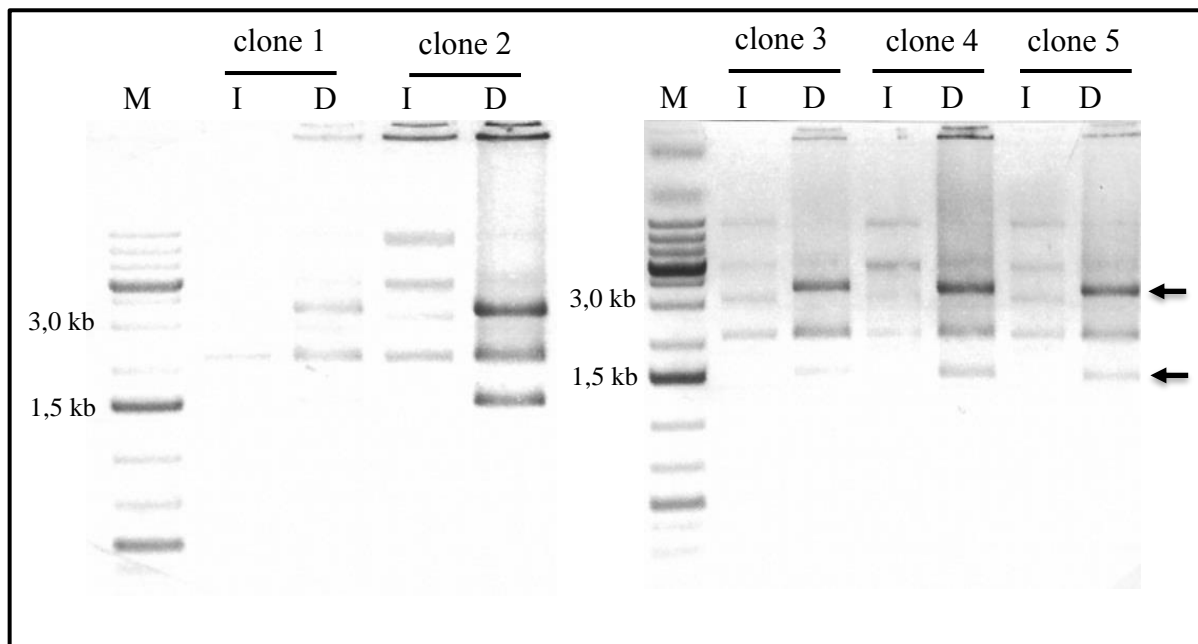
**Figura 10. Estratégia para a construção do vetor pKId-GFP.** O vetor final contém a marca *leu2-d* e o gene repórter *EGFP* sob controle do promotor *PGK1*.

Inicialmente, o alelo *leu2-d* foi obtido por PCR a partir do DNA genômico de *S. cerevisiae* usando os primers 5-leud e 3-leud (tabela 2). O fragmento amplificado de ~1,4 kb corresponde à região codante do gene *LEU2* com sua região terminadora da transcrição e apenas 29 pb do promotor. O fragmento foi clonado no vetor pPICK2 linearizado com BamHI gerando o vetor pKId. Células de *E. coli* XL10-gold foram transformadas e a seleção se deu em meio LB contendo kanamicina.

A estratégia usada para deletar o gene *LEU2* na linhagem M12 (Betancur *et al.*, 2017) deixou a região promotora intacta, portanto, uma única cópia desse gene integrada no locus *leu2* poderia restabelecer prototrofia à levedura. Por esse motivo, neste trabalho foi usado o

alelo *leu2-d* de *S. cerevisiae* ao invés da versão de *K. phaffii* para evitar a integração do vetor no locus *leu2* por recombinação homóloga.

Apesar de ter sido parcial, a digestão com KpnI e NotI permitiu confirmar a clonagem do alelo *leu2-d* nos cinco clones analisados, assim como determinar sua orientação no vetor. Na figura 11 pode ser observado que a digestão resultou em fragmentos de ~1,5 kb e ~3,5 kb que eram esperados caso o fragmento tivesse entrado no sentido anti-horário.

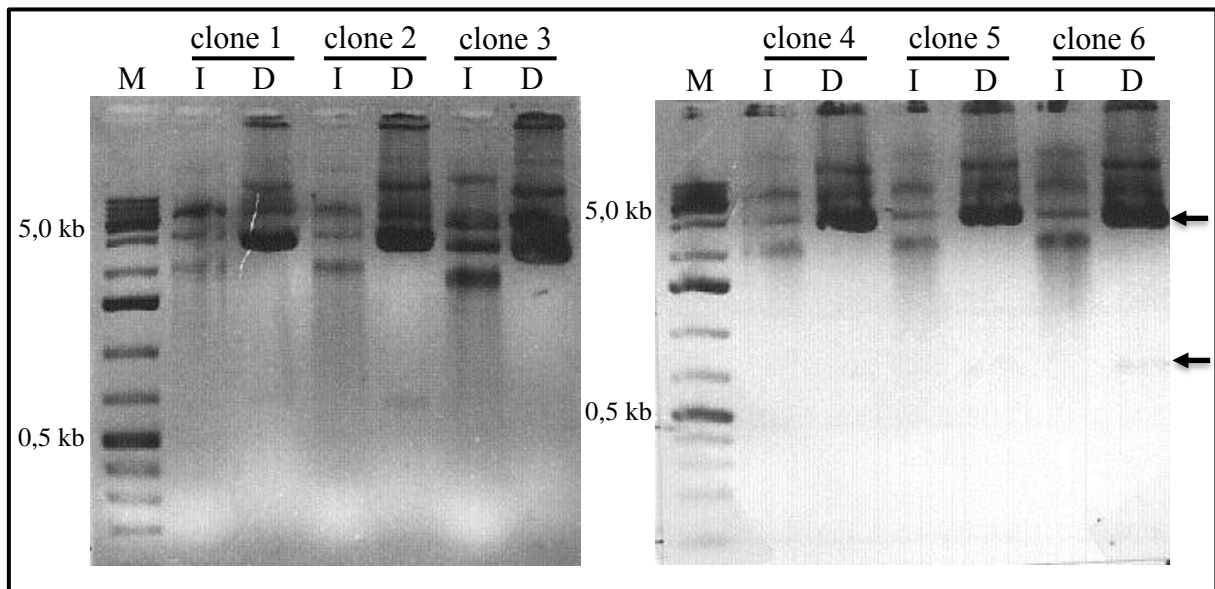


**Figura 11. Clonagem do alelo *leu2-d* no vetor pPICK2.** Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando o DNA plasmidial intacto (I) e digerido com KpnI e NotI (D) para cinco clones transformantes. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). As setas indicam os fragmentos esperados.

Já foi observado que os promotores presentes no plasmídeo podem ter uma atividade de transcrição na marca defectiva. Por exemplo, as linhagens auxotróficas *ade1* e *ade2* de *K. phaffii* foram capazes de crescer em meio mínimo quando transformadas com um plasmídeo contendo apenas a região codante (sem promotor) dos genes *ADE1* e *ADE2* (Du *et al.*, 2012). Por este motivo, a marca *leu2-d* foi clonada na direção oposta aos promotores presentes no vetor (*PGK1* e *TEF1*) (Figura 10).

O vetor pKld foi digerido com SacI e NotI para remover o sinal de secreção do fator  $\alpha$ . A digestão com essas enzimas removeu também um fragmento de ~200 pb do promotor

*PGK1* o qual foi restituído quando o gene EGFP foi subclonado. O fragmento de ~1 kb contendo o gene EGFP fusionado com o pedaço faltante do promotor *PGK1* foi obtido pela digestão do vetor pPICK-GFP (vetor previamente desenvolvido no nosso laboratório e derivado do vetor pPIC9 contendo o gene EGFP sob controle do promotor *PGK1*) com as enzimas SacI e NotI. A ligação desse fragmento no pKld gerou o vetor pKld-GFP. Células de *E. coli* XL10-gold foram transformadas e selecionadas em meio LB contendo kanamicina. A clonagem foi confirmada por digestão com BglII e BamHI que deveria resultar em fragmentos de ~600 pb e ~4,9 kb, caso o gene EGFP tivesse sido ligado no vetor. Como observado na figura 12, os clones 2 e 6 apresentaram bandas dos tamanhos esperados. Um clone foi selecionado e o vetor obtido foi chamado de pKld-GFP.



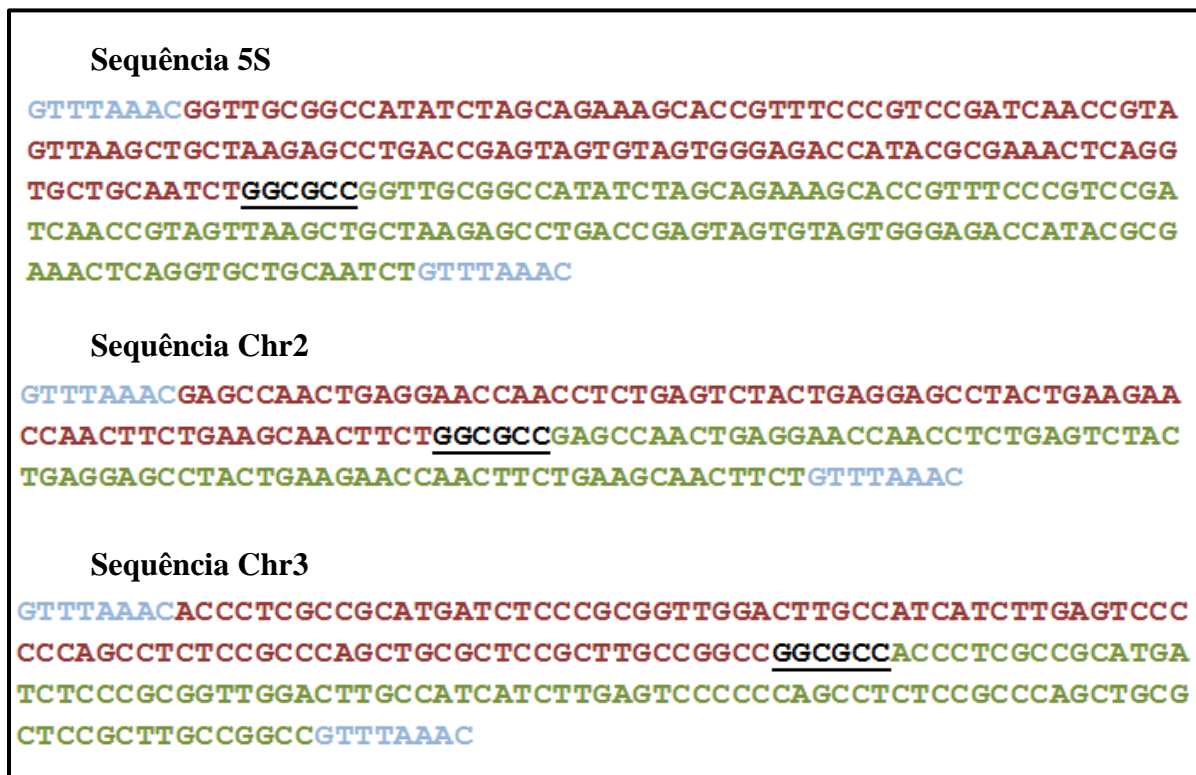
**Figura 12. Clonagem do gene EGFP no vetor pKld.** Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando o DNA plasmidial intacto (I) e digerido com BglII e BamHI (D) para seis clones transformantes. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). As setas indicam os fragmentos esperados.

### 5.2.2 Construção dos vetores pBSChr2ld-GFP, pBSChr3ld-GFP, pBS5Sld-GFP e pPCVNTSld-GFP

As sequências selecionadas dos cromossomos 2 (Chr2) e 3 (Chr3) e a sequência 5S foram sintetizadas de maneira que duas cópias das mesmas ficassem separadas por um sítio de

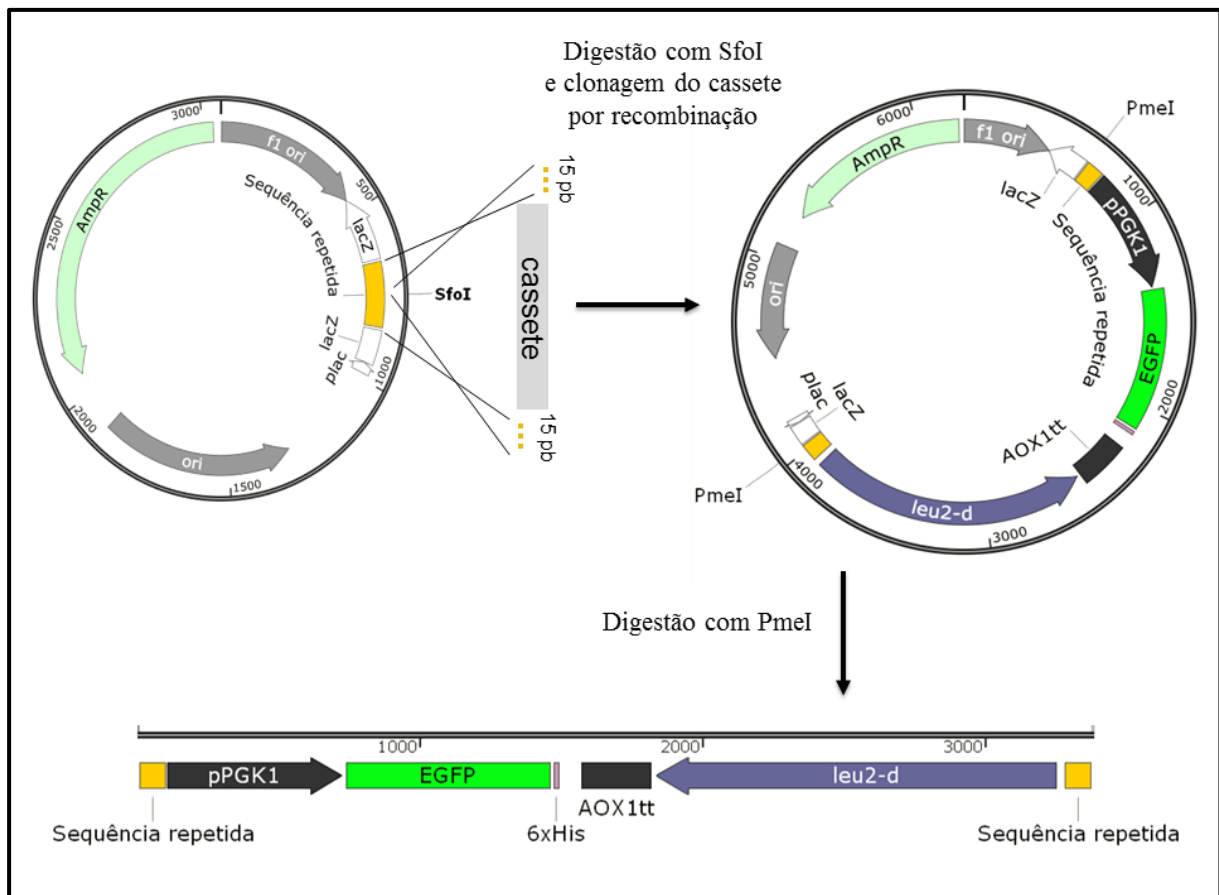


restrição para a enzima SfoI. Nas extremidades das sequências sintéticas foram adicionados sítios para a enzima PmeI, como mostrado na figura 13.



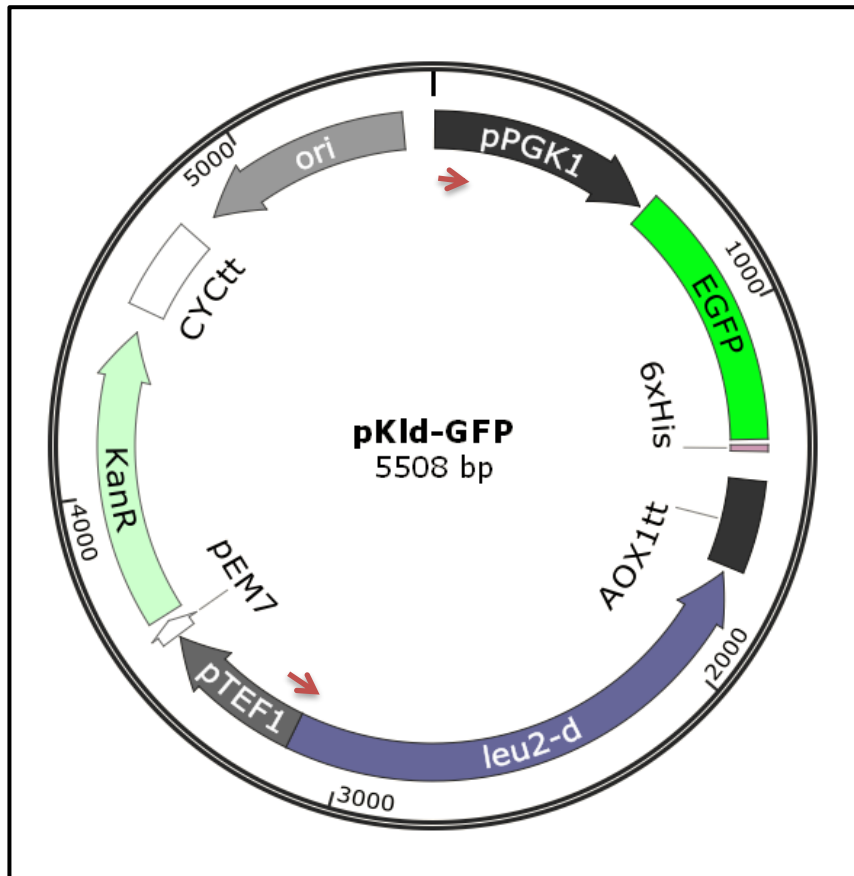
**Figura 13. Sequências repetidas sintetizadas.** Em azul, sítios para PmeI; em preto sublinhado, sítios para SfoI.

Para direcionar a integração às regiões repetidas no genoma de *K. phaffii* foram construídos cassetes de expressão flanqueados por essas sequências para permitir a recombinação homóloga nesses *loci*. Todos os cassetes têm em comum um fragmento de ~3 kb formado pelo promotor *PGK1*, o gene EGFP, o terminador da transcrição do gene *AOX1* e a marca *leu2-d*, flanqueado pelas sequências repetidas. Os sítios para a enzima PmeI presentes nas extremidades das sequências repetidas permitem liberar o cassete de expressão. A figura 14 apresenta a estratégia usada para a construção e liberação desse cassete.



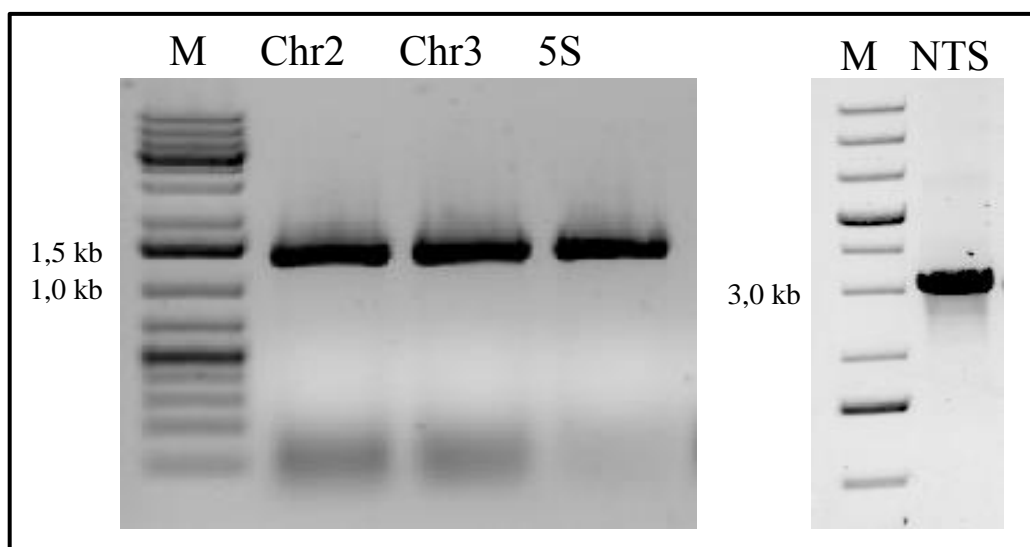
**Figura 14. Estratégia para a construção de cassete de expressão com seqüências repetitivas.** Cassete para expressar o gene EGFP sob controle do promotor *PGK1* e com o alelo *leu2-d* como marca de seleção. As seqüências repetidas flanqueadoras direcionam a integração do cassete.

O fragmento PGK-EGFP-leu2d foi obtido por PCR com pares de *primers* que contêm 15 pb homólogos a cada seqüência repetida para realizar a clonagem por recombinação homóloga usando o kit In-fusion (*primers* PGK-*chr2* e leud-*chr2*, PGK-*chr3* e leud-*chr3*, PGK-5S e leud-5S, PGK-NTS e leud-NTS, tabela 2). Como molde foi usado o vetor pKId-GFP. A figura 15 mostra os sítios onde se anelam os diferentes *primers* PGK e leud.



**Figura 15. Sítios de anelamento dos *primers* para amplificar o cassete PGK-EGFP-leud.** As setas vermelhas indicam a região onde anelam os diferentes *primers* PGK e leud. Tamanho do amplicon: ~3 kb.

Cada uma das sequências sintéticas foi clonada no pBlueScript SK II seguido de digestão com *SfoI* para linearizar o plasmídeo. Os fragmentos PGK-GFP-leu2d com sequências homólogas às sequências sintéticas foram clonados no respectivo plasmídeo usando o kit In-fusion. Células competentes de *E. coli* HST08 (*Stellar Competent Cells*) foram transformadas com os três sistemas e a seleção de transformantes foi feita em meio LB contendo ampicilina. Para confirmar a clonagem foi feita uma PCR usando como *primer forward* os respectivos *primers* PGK (PGK-*chr2*, PGK-*chr3* ou PGK-5S) e, como *primer reverse*, o primer EGFP-Not (tabela 2) que se anela na região 3' do gene EGFP. Como pode ser observado na figura 16, com os três vetores houve amplificação de um fragmento de ~1,4 kb, como era esperado, confirmando a presença do fragmento PGK-GFP-leu2d. Os vetores obtidos contendo a sequência dos cromossomos 2, 3 e a sequência 5S foram chamados de pBSChr2ld-GFP, pBSChr3ld-GFP e pBS5Sld-GFP, respectivamente.

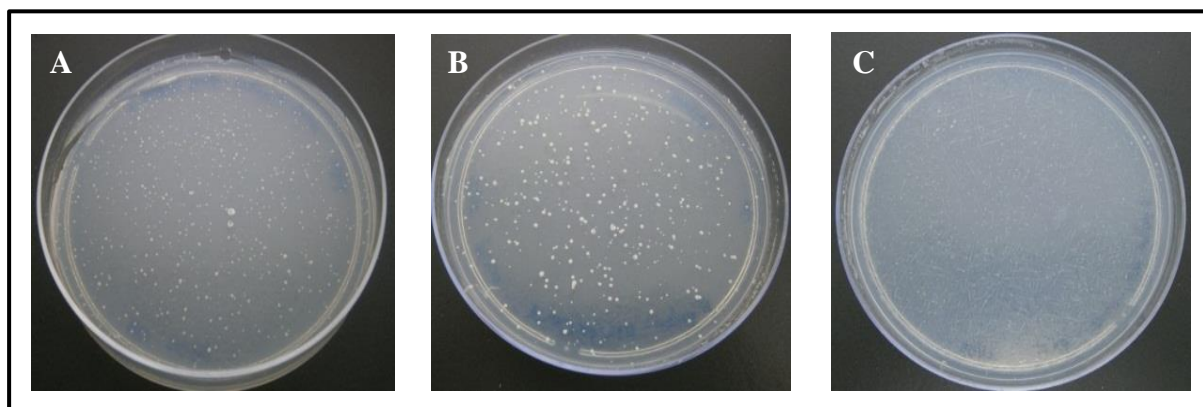


**Figura 16. Confirmação da clonagem do fragmento PGK-EGFP-leu2d nos plasmídeos com sequências repetidas.** Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando o resultado da PCR. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific), Chr2: reação com o vetor pBSChr2ld-GFP como template, Chr3: reação com o vetor pBSChr3ld-GFP como template, 5S: reação com o vetor pBS5Sld-GFP como template, NTS: reação com o vetor pPCVNTSld-GFP como template.

A sequência NTS do rDNA foi obtida por PCR com os *primers* 18SF-pPCV e 25SR-pPCV a partir do DNA genômico de *K. phaffii* e o fragmento amplificado foi clonado no plasmídeo pPCV (Janner *et al.*, 2013) linearizado com EcoRV. A digestão com a enzima SpeI linearizou o plasmídeo no meio da sequência NTS e, nessa região, foi clonado o fragmento PGK-EGFP-leu2d com o kit In-fusion. Células competentes de *E. coli* XL10-gold foram transformadas com o sistema e a seleção se deu em meio LB contendo ampicilina. Para confirmar a clonagem, foi feita uma PCR usando os *primers* 25SR-pPCV e 3-leud (tabela 2). Na figura 16 é possível observar que houve amplificação de um fragmento de ~3 kb como era esperado, confirmando a presença do fragmento PGK-EGFP-leu2d. O vetor obtido foi chamado de pPCVNTSld-GFP.

### 5.2.3 Transformação de *K. phaffii* M12

*K. phaffii* M12 foi transformada com o vetor pKld-GFP linearizado com SacI e com os quatro cassetes contendo as sequências repetidas liberados após digestão dos respectivos vetores com PmeI. Em todos os casos, a seleção de transformantes foi feita em meio MD já que com a presença da marca *leu2-d* a levedura deveria recuperar a prototrofia. Quando a levedura foi transformada com o vetor pKld-GFP só começaram a aparecer colônias nas placas após quatro dias de incubação e foi observado que algumas colônias cresciam mais rápido que outras. A figura 17 mostra o crescimento das colônias transformantes após sete dias de incubação. Na placa controle, onde foi plaqueada a levedura submetida a eletroporação sem DNA, não houve crescimento.

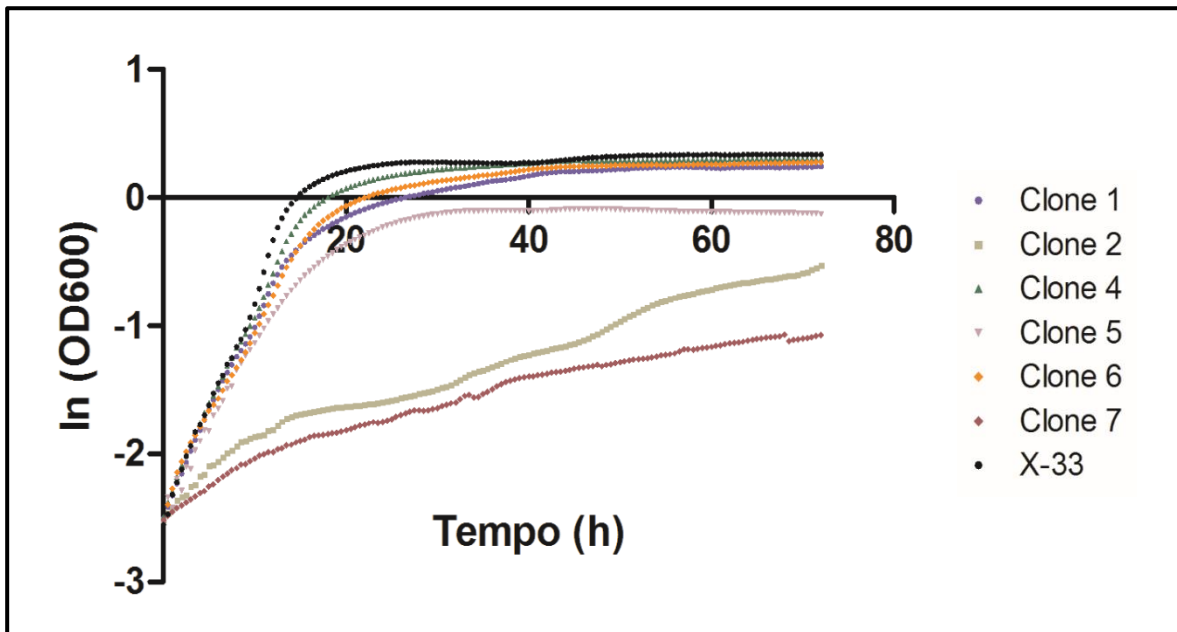


**Figura 17. Resultado da transformação de *K. phaffii* M12 com o vetor pKld-GFP.** Placas com meio MD após 7 dias de incubação a 30°C. **A e B** células transformadas com o vetor, **C** células submetidas a eletroporação sem DNA (controle negativo).

Dez clones transformantes foram selecionados para análise, porém, os clones 3, 8, 9 e 10 foram descartados já que depois de várias gerações não foram capazes de crescer em meio mínimo. Provavelmente, esses falsos positivos integraram apenas uma cópia do vetor ou perderam cópias do cassete de integração por recombinação intramolecular ficando com um número de cópias insuficiente para sustentar prototrofia.

Para avaliar se havia diferença na velocidade de crescimento dos seis clones restantes foi feita uma cinética de crescimento em meio MD. O crescimento foi comparado com o da levedura selvagem X-33. Como observado na figura 18, houve clones que apresentaram um perfil de crescimento parecido ao de X-33 (clones 1, 4 e 6), enquanto os clones 2 e 7 cresceram muito mais lentamente. A tabela 7 apresenta a velocidade específica máxima de

crescimento ( $\mu_{\text{máx}}$ ) para cada um dos clones avaliados. A análise estatística mostrou que não existe diferença significativa entre a velocidade específica máxima de crescimento dos clones 1, 4, 5 e 6 em comparação a X-33, enquanto que a diferença é significativa para os clones 2 e 7 quando comparados com a linhagem selvagem. Os clones com maiores taxas de crescimento foram os que apareceram primeiro na placa após a transformação.



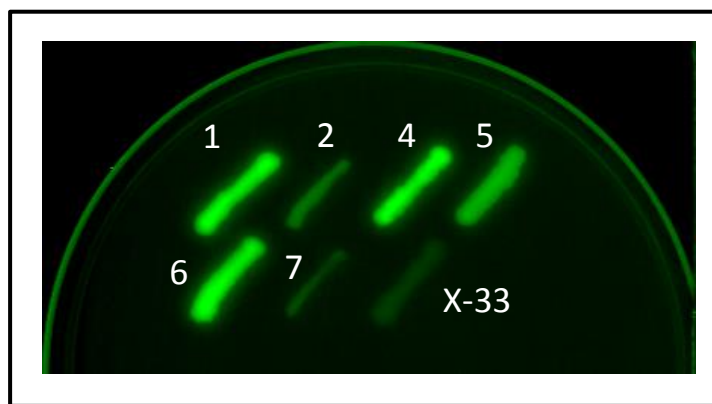
**Figura 18. Cinética de crescimento dos clones transformados com o vetor pKId-GFP.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,08. X-33: linhagem selvagem de *K. phaffii*.

**Tabela 7.** Velocidade específica máxima de crescimento dos clones transformados com o vetor pKId-GFP

Clone	$\mu_{\text{máx}}$ (h <sup>-1</sup> )
1	0,1596 ± 0,0062
2	0,0708 ± 0,0022*
4	0,1614 ± 0,0027
5	0,1408 ± 0,0075
6	0,1650 ± 0,0078
7	0,0502 ± 0,0023*
X-33	0,1707 ± 0,0069

\*Diferença significativa em comparação à linhagem selvagem X-33 ( $p < 0,05$ ). Análise de variância (ANOVA) seguida do teste de Tukey.

Os seis clones transformantes foram semeados em meio MD e a placa foi analisada no Typhoon Scanner para avaliar a produção de GFP. Como observado na figura 19, todos os clones apresentaram fluorescência, o que confirmou a integração do vetor pKld-GFP e a expressão do gene EGFP nas células transformantes. A intensidade de fluorescência apresentada pelos clones 2 e 7 foi muito fraca, quase igual à que apresentou a levedura selvagem (*K. phaffii* X-33). Já os clones 1, 4, 5 e 6 mostraram uma intensidade de fluorescência maior. Esses clones que exibiram maior fluorescência foram os que apresentaram maiores taxas de crescimento, o que poderia estar relacionado com a integração de um maior número de cópias do vetor de expressão.

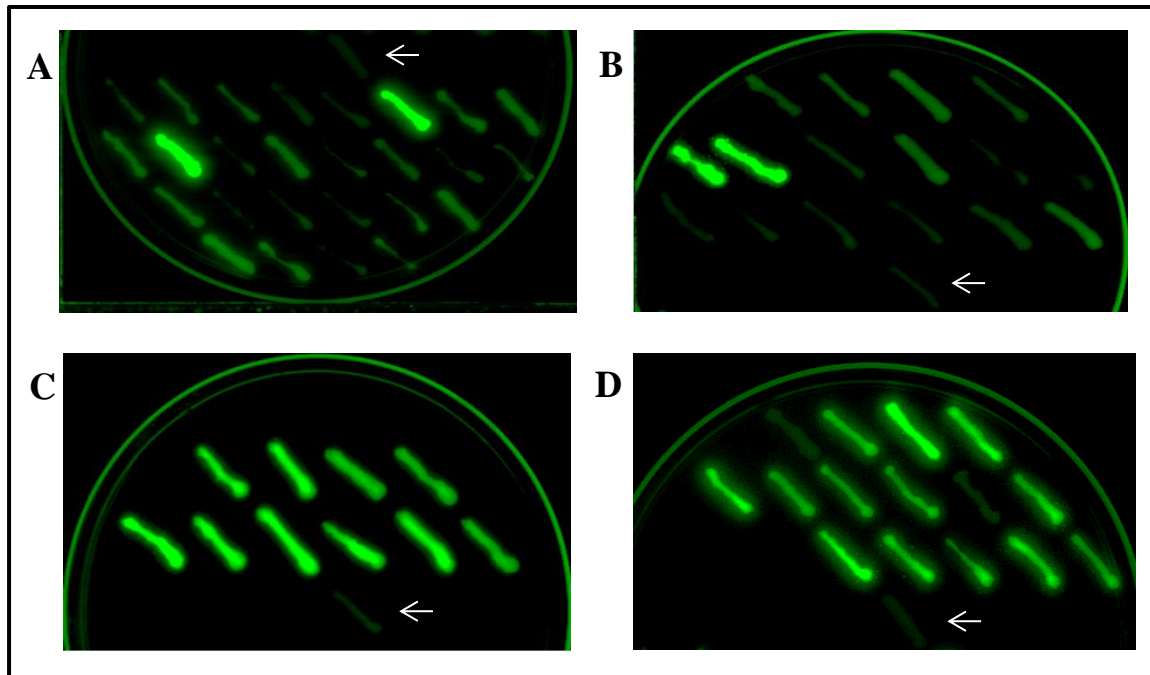


**Figura 19. Expressão intracelular de EGFP em clones transformados com o vetor pKld-GFP.** Placa com meio MD incubada por 48 h a 30°C e analisada no Typhoon Scanner.

Quando *K. phaffii* M12 foi transformada com os quatro cassetes contendo as sequências repetidas foi observado um lento crescimento das células, sendo que as colônias começaram a aparecer depois de sete dias de incubação e, semelhantemente ao que aconteceu com o vetor pKld-GFP, houve colônias com diferentes tamanhos.

Vinte e seis clones transformados com o cassete Chr2ld-GFP, 16 transformados com o cassete Chr3ld-GFP, 10 transformados com o cassete 5Sld-GFP e 15 transformados com o cassete NTSld-GFP foram selecionados para analisar a produção intracelular de GFP no Typhoon Scanner. Como observado na figura 20, todos os clones transformantes apresentaram fluorescência e foi possível identificar alguns com maior intensidade que o controle *K. phaffii* X-33. No caso das células transformadas com os cassetes Chr2ld-GFP e Chr3ld-GFP, foram observados diferentes níveis de intensidade de fluorescência entre os

clones analisados, enquanto que as células transformadas com os cassetes 5Sld-GFP e NTSld-GFP mostraram níveis de intensidade mais homogêneos.

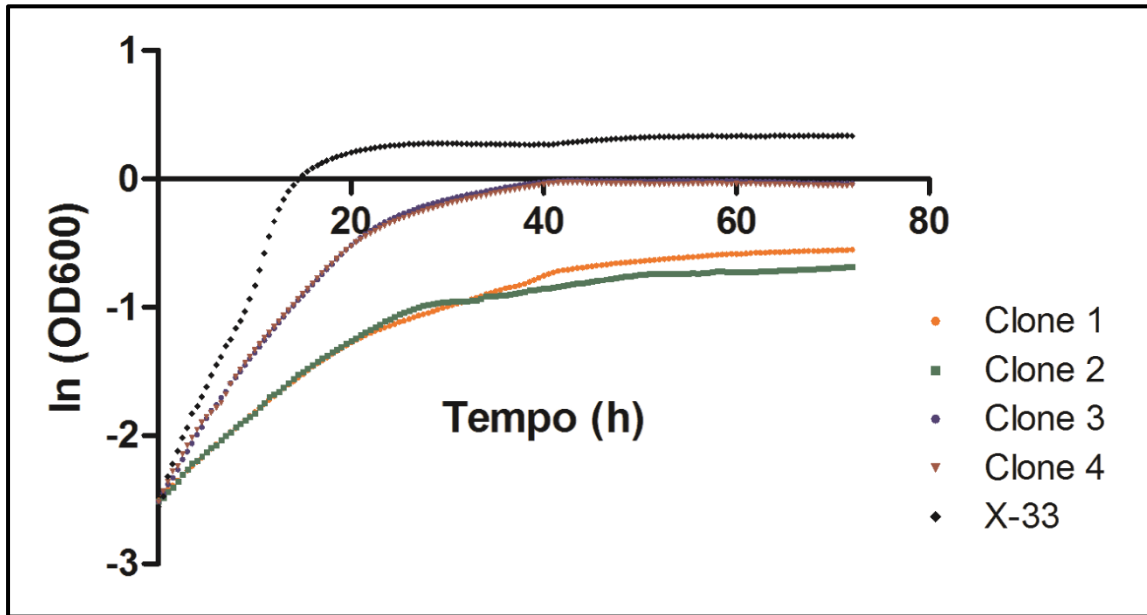


**Figura 20. Expressão intracelular de EGFP em clones transformados com os cassetes contendo seqüências repetidas.** Placa com meio MD incubada por 48 h a 30°C e analisada no Typhoon Scanner. **A** Clones transformados com Chr2ld-GFP, **B** Clones transformados com Chr3ld-GFP, **C** Clones transformados com 5Sld-GFP, **D** Clones transformados com NTSld-GFP. A seta indica o controle *K. phaffii* X-33.

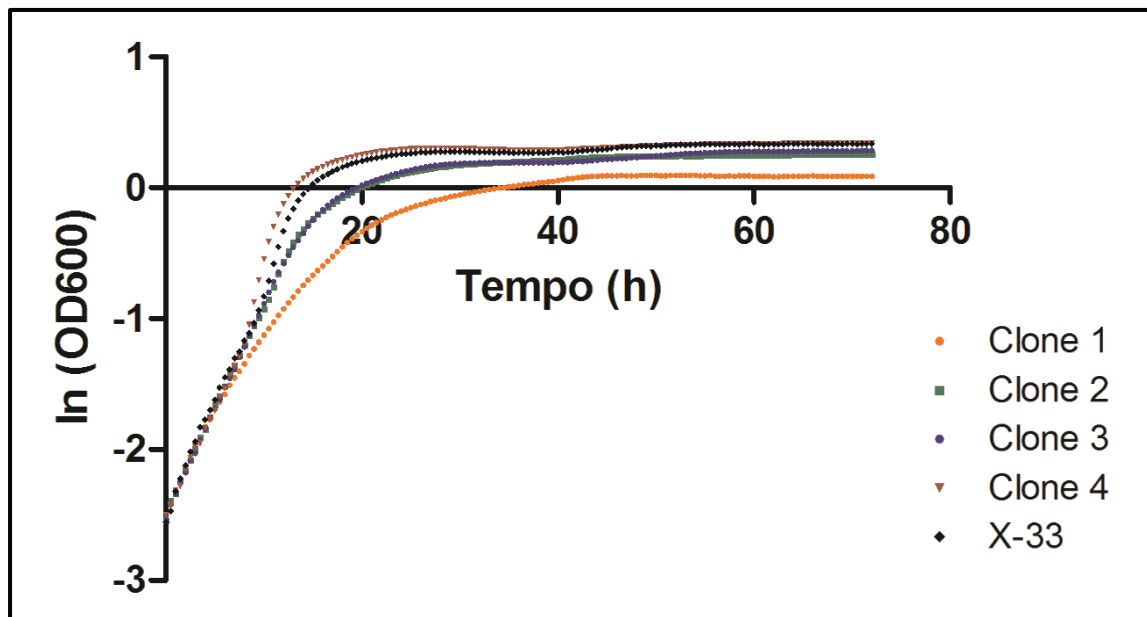
Quatro clones de cada sistema foram selecionados para posteriores análises. No caso dos cassetes Chr2ld-GFP e Chr3ld-GFP foram escolhidos os dois clones que apresentaram maior fluorescência e mais dois clones aleatórios. Para os outros dois cassetes a seleção foi completamente aleatória.

Foi feita uma cinética de crescimento em meio MD dos quatro clones selecionados para cada sistema e o crescimento foi comparado com o da levedura selvagem X-33. As figuras 21, 22, 23 e 24 apresentam as curvas de crescimento para os clones transformados com os cassetes Chr2ld-GFP, Chr3ld-GFP, 5Sld-GFP e NTSld-GFP, respectivamente. A tabela 8 apresenta a velocidade específica máxima de crescimento ( $\mu_{\text{máx}}$ ) determinada para cada um dos clones avaliados dos quatro sistemas.

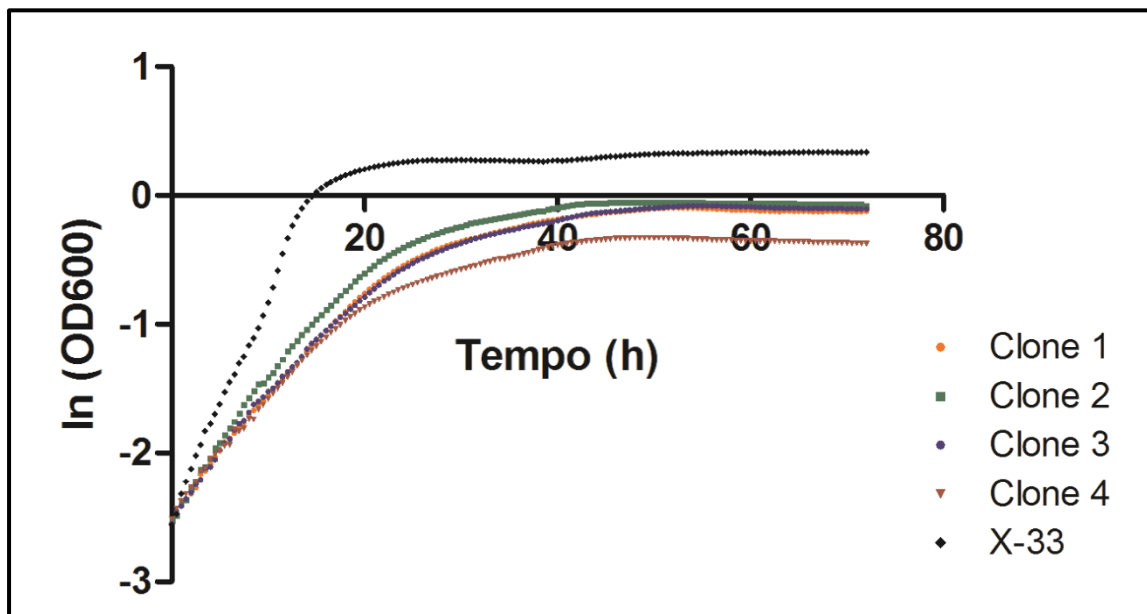




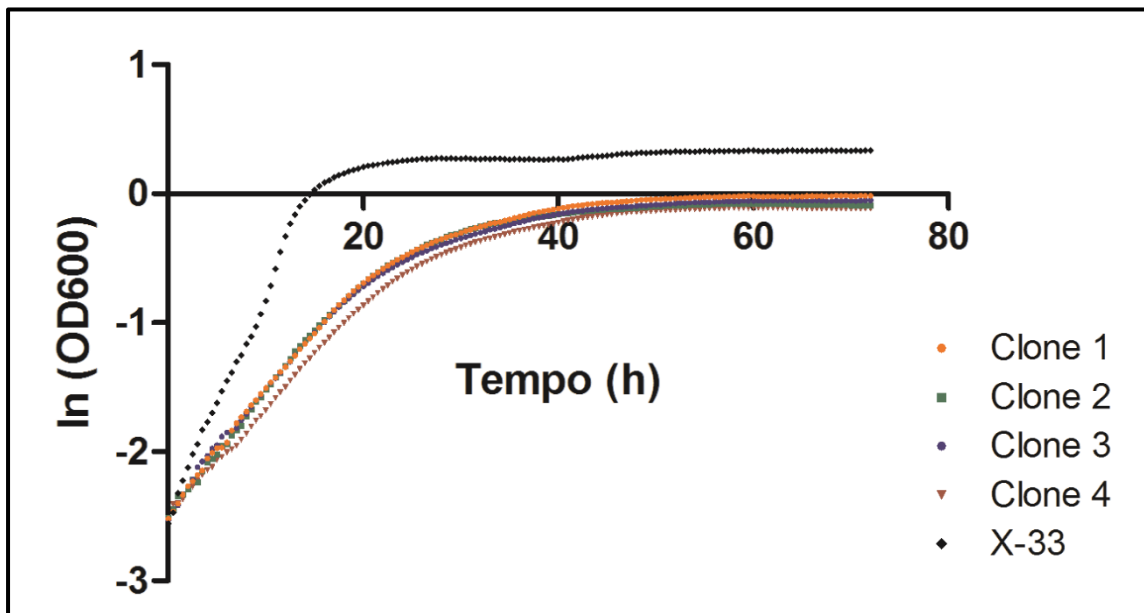
**Figura 21. Cinética de crescimento dos clones transformados com o cassete Chr2ld-GFP.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,08. X-33: linhagem selvagem de *K. phaffii*.



**Figura 22. Cinética de crescimento dos clones transformados com o cassete Chr3ld-GFP.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,08. X-33: linhagem selvagem de *K. phaffii*.



**Figura 23. Cinética de crescimento dos clones transformados com o cassete 5SId-GFP.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,08. X-33: linhagem selvagem de *K. phaffii*.



**Figura 24. Cinética de crescimento dos clones transformados com o cassete NTSId-GFP.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,08. X-33: linhagem selvagem de *K. phaffii*.

**Tabela 8.** Velocidade específica máxima de crescimento dos clones transformados com os cassetes Chr2ld-GFP, Chr3ld-GFP, 5Sld-GFP e NTSld-GFP.

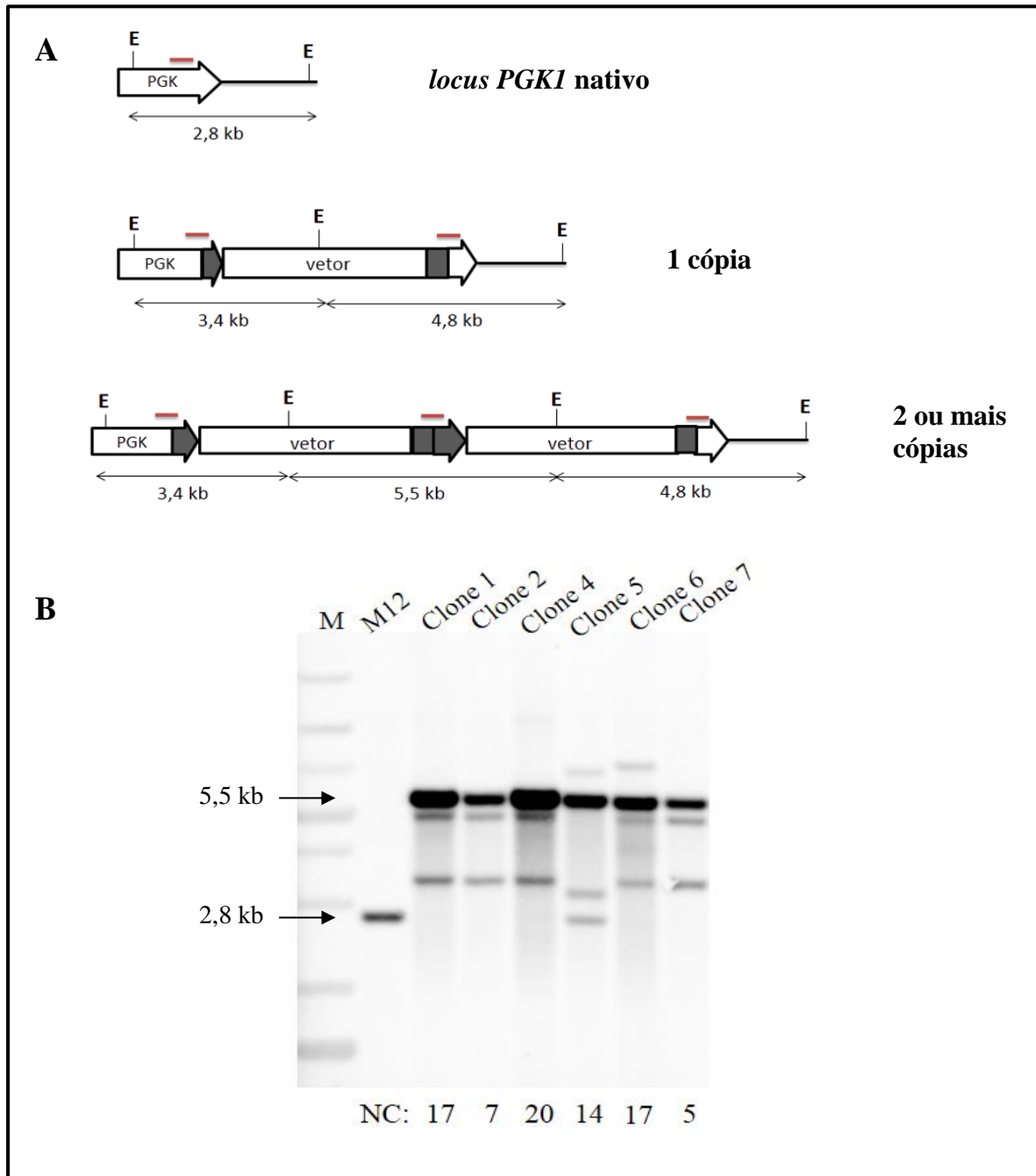
Sistema	Clone	$\mu_{\text{máx}} (\text{h}^{-1})$
Chr2ld-GFP	1	$0,0703 \pm 0,0007^*$
	2	$0,0725 \pm 0,0012^*$
	3	$0,1225 \pm 0,0020^*$
	4	$0,1244 \pm 0,0035^*$
Chr3ld-GFP	1	$0,1521 \pm 0,0036$
	2	$0,1650 \pm 0,0028$
	3	$0,1620 \pm 0,0040$
	4	$0,1621 \pm 0,0053$
5Sld-GFP	1	$0,1026 \pm 0,0028^*$
	2	$0,1198 \pm 0,0022^*$
	3	$0,1052 \pm 0,0018^*$
	4	$0,0928 \pm 0,0026^*$
NTSld-GFP	1	$0,0822 \pm 0,0025^*$
	2	$0,0920 \pm 0,0026^*$
	3	$0,0888 \pm 0,0008^*$
	4	$0,0909 \pm 0,0024^*$
X-33		$0,1707 \pm 0,0069$

\*Diferença significativa em comparação à linhagem selvagem X-33 ( $p < 0,05$ ). Análise de variância (ANOVA) seguida do teste de Tukey.

Apenas os clones transformados com o cassete Chr3Id-GFP apresentaram um perfil de crescimento e uma velocidade específica máxima de crescimento semelhante ao da levedura selvagem. A velocidade específica máxima de crescimento foi menor para todos os clones dos outros três sistemas em comparação a X-33, sendo que, entre os quatro clones dos sistemas Chr3Id-GFP, 5SId-GFP e NTSId-GFP, não houve diferença significativa. No caso dos clones transformados com o cassete Chr2Id-GFP, a velocidade específica máxima de crescimento dos clones 1 e 2 foi significativamente menor que a velocidade dos clones 3 e 4.

#### 5.2.4 Determinação do número de cópias

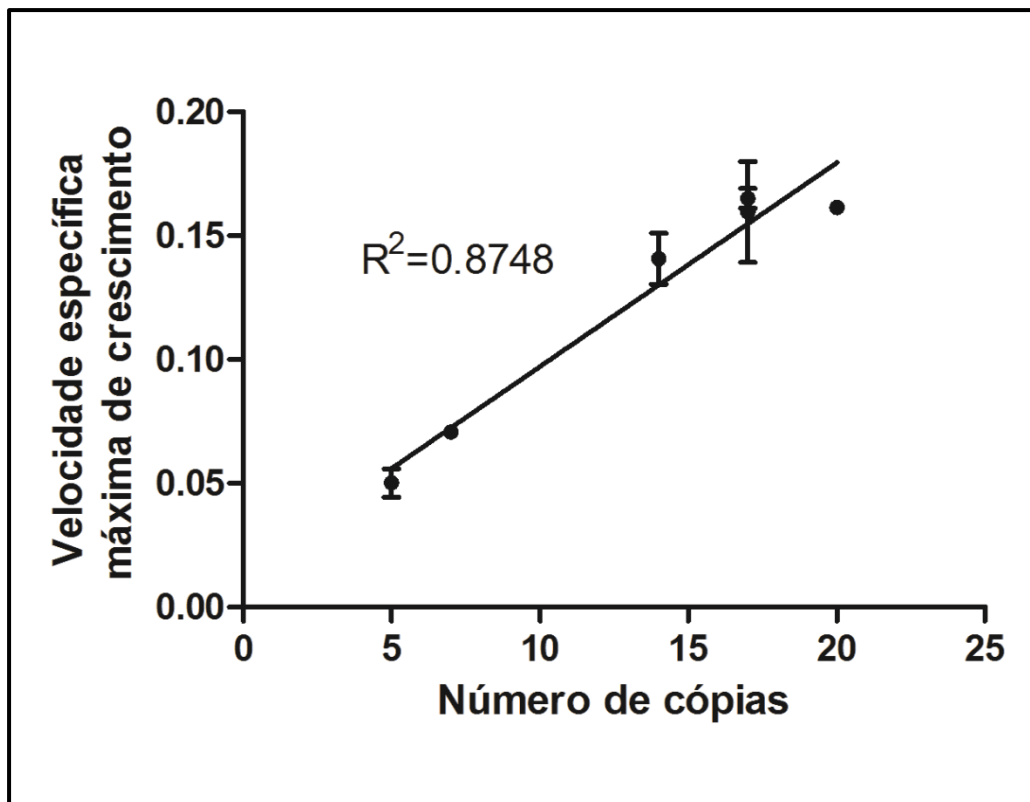
O DNA dos seis clones transformados com o vetor pKId-GFP foi digerido com EcoRI e submetido a análise por Southern blot para determinação do número de cópias integradas do vetor. A sonda utilizada, que hibridiza no promotor *PGK1*, devia permitir a identificação de dois fragmentos de ~3,4 kb e ~4,8 kb se apenas uma cópia do vetor tivesse sido integrada. Se tivessem sido integradas duas cópias do vetor, seria identificado um terceiro fragmento de ~5,5 kb que corresponde ao tamanho do vetor e a intensidade desse fragmento iria aumentando por cada cópia a mais integrada. Um fragmento de ~2,8 kb correspondente ao *locus* do *PGK1* seria observado se a integração ocorresse em outro *locus* diferente. A figura 25A apresenta um esquema dos fragmentos esperados e a figura 25B mostra o resultado do Southern blot. Como pode ser observado, todos os clones apresentaram o fragmento de ~5,5 kb, confirmando a integração de várias cópias do vetor. O fragmento de ~2,8 kb observado na linhagem M12 correspondente ao *locus PGK1* intacto que também foi observado no clone 5, indicando que este clone apresentou múltiplas cópias do vetor integradas em outro local. Todos os demais clones tiveram o vetor integrado *in tandem* no *locus PGK1* como era esperado caso houvesse recombinação homóloga.



**Figura 25. Determinação do número de cópias dos clones transformados com o vetor pKld-GFP.** A) Esquema dos fragmentos esperados após digestão do gDNA com a enzima EcoRI, B) Resultado do Southern blot. E: sítio de corte para a enzima EcoRI. NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). A linha vermelha indica o lugar onde hibridiza a sonda.

O número de cópias integradas foi determinado após comparação da intensidade da banda de ~5,5 kb em relação à banda de ~3,4 kb (ou à banda de ~2,8 kb, no caso do clone 5)

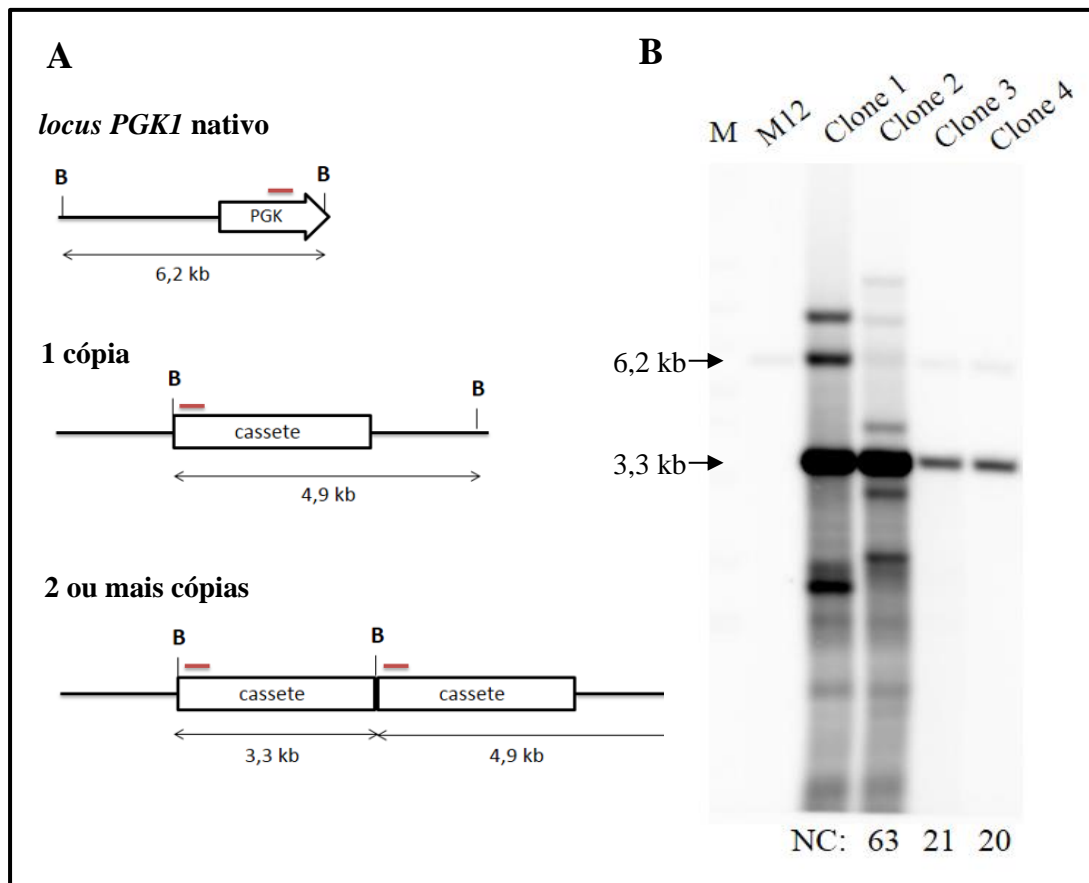
que tem uma única cópia. Os clones 1, 4, 5 e 6, que apresentaram maiores taxas de crescimento, foram os que apresentaram o maior número de cópias integradas do vetor: 17, 20, 14 e 17, respectivamente. Os clones 2 e 7, com crescimento lento, apresentaram 7 e 5 cópias, respectivamente. Esses resultados mostraram uma relação direta entre o aumento da taxa de crescimento com o número de cópias integradas com uma correlação linear ( $R^2 = 0.8748$ ) como pode ser observado na figura 26. Esta observação era esperada já que os clones que apresentaram rápido crescimento deviam ter um maior número de cópias da marca *leu2-d* integradas para complementar a auxotrofia. Além disso, o resultado está em concordância com o obtido quando os genes *ADE1* e *ADE2* truncados foram usados como marcas de seleção defectivas para *K. phaffii* e foram identificados clones multicópia como colônias de tamanho maior (Du *et al.*, 2012).



**Figura 26.** Correlação entre a velocidade específica de crescimento e o número de cópias dos clones transformados com o vetor pKld-GFP. As barras de erro representam o erro padrão da média para  $n=3$ .

O DNA dos quatro clones analisados transformados com o cassete Chr2ld-GFP foi digerido com BglII. A sonda utilizada que hibridiza no promotor *PGK1* deveria permitir a identificação de um fragmento de ~4,9 kb se apenas uma cópia do vetor tivesse sido

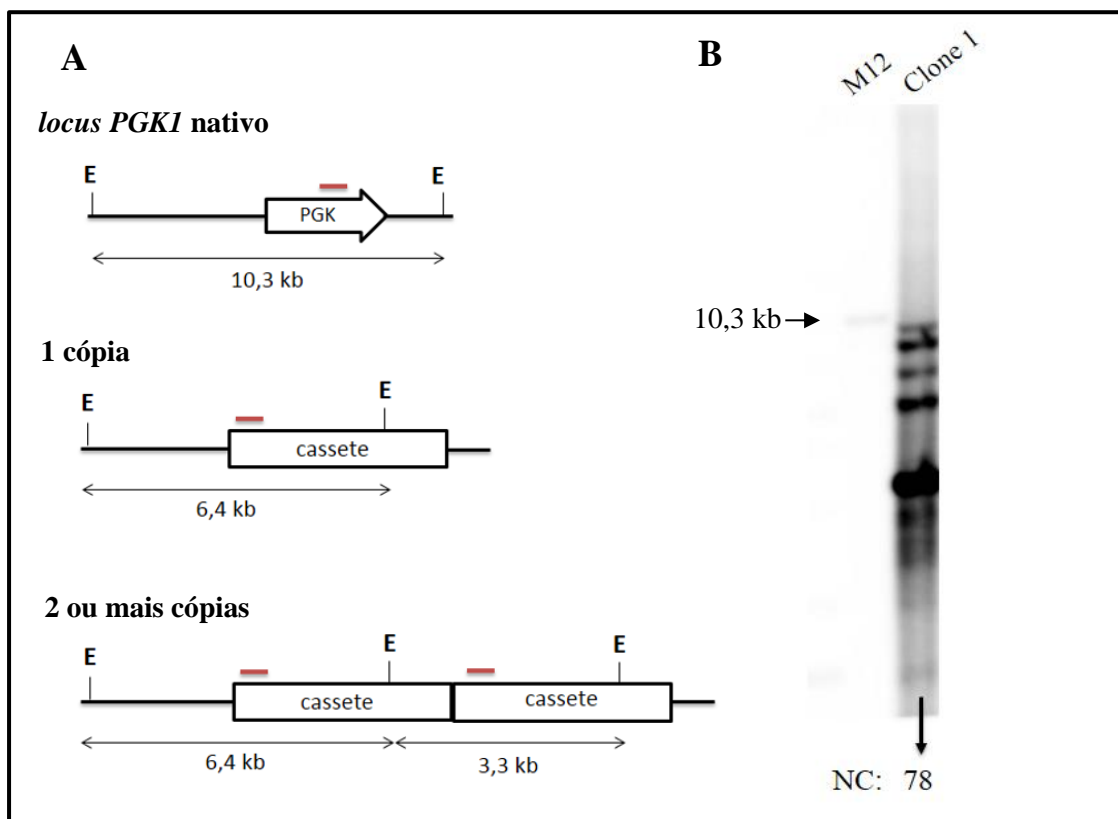
integrada. Se duas cópias do vetor tivessem sido integradas seria identificado um outro fragmento de ~3,3 kb, que corresponde ao tamanho do cassete e a intensidade desse fragmento iria aumentando por cada cópia a mais integrada. Um fragmento de ~6,2 kb correspondente ao *locus* do *PGK1* seria observado se a integração ocorresse em outro *locus* diferente. A figura 27A apresenta um esquema dos fragmentos esperados e a figura 27B mostra o resultado do Southern blot.



**Figura 27. Determinação do número de cópias dos clones transformados com o cassete Chr2Id-GFP.** A Esquema dos fragmentos esperados após digestão do gDNA com a enzima BglII, B Resultado do Southern blot. B: sítio de corte para a enzima BglII. NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). A linha vermelha indica o lugar onde hibridiza a sonda.

Os clones 1 e 2 apresentaram mais fragmentos do que esperado o que indicaria integração em outros locais diferentes, porém, o fragmento de ~3,3 kb, correspondente ao tamanho do cassete, foi identificado em todos os clones confirmando integração *in tandem*. Para determinar o número de cópias, foi comparada a intensidade da banda de ~3,3 kb com a intensidade da banda de ~6,2 kb, que corresponde a uma cópia no genoma da levedura e que

foi observada em todos os clones confirmando que a integração não ocorreu no *locus PGK1*. Esse fragmento controle apareceu com intensidade mais forte no clone 1, sendo que a quantidade de DNA digerido foi similar para os quatro clones. Essa observação indicou que poderia existir um outro fragmento do mesmo tamanho que o fragmento correspondente ao *PGK1* intacto e, portanto, não podia ser usado como referência para determinar o número de cópias integradas no clone 1. A digestão do DNA do clone 1 com a enzima EcoRV demonstrou que houve integração no *locus PGK1* já que não foi observado o fragmento de ~10,3 kb correspondente a esse *locus* intacto (figura 28).



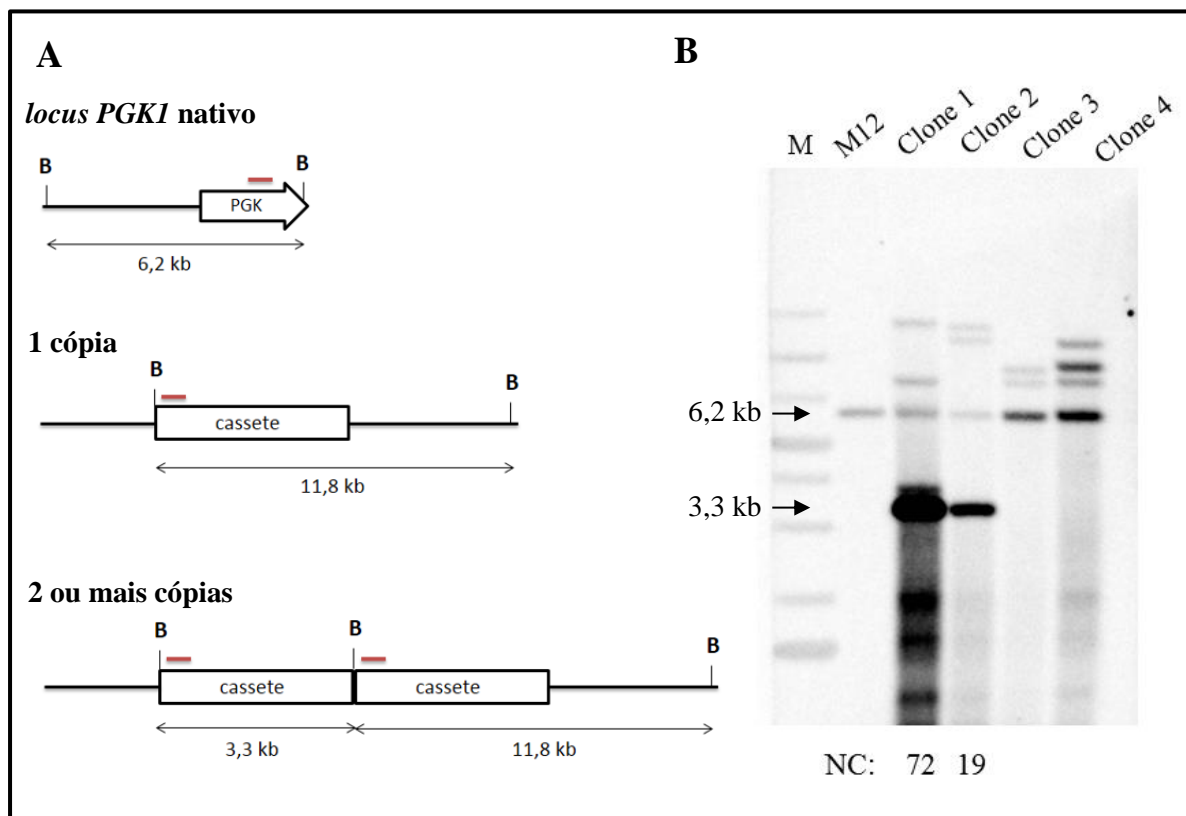
**Figura 28. Determinação do número de cópias do clone 1 transformado com o cassette Chr2ld-GFP.** **A** Esquema dos fragmentos esperados após digestão do gDNA com a enzima EcoRV, **B** Resultado do Southern blot. E: sítio de corte para a enzima EcoRV. NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). A linha vermelha indica o lugar onde hibridiza a sonda.

Dessa forma, o número de cópias para o clone 1 foi determinado usando como referência o fragmento de ~6,2 kb do clone 2. Os quatro clones analisados integraram um alto número de cópias, sendo que os clones 1 e 2 integraram uma quantidade muito maior de cópias (78 e 63 cópias, respectivamente), porém, em outros *loci* além daquele esperado. No



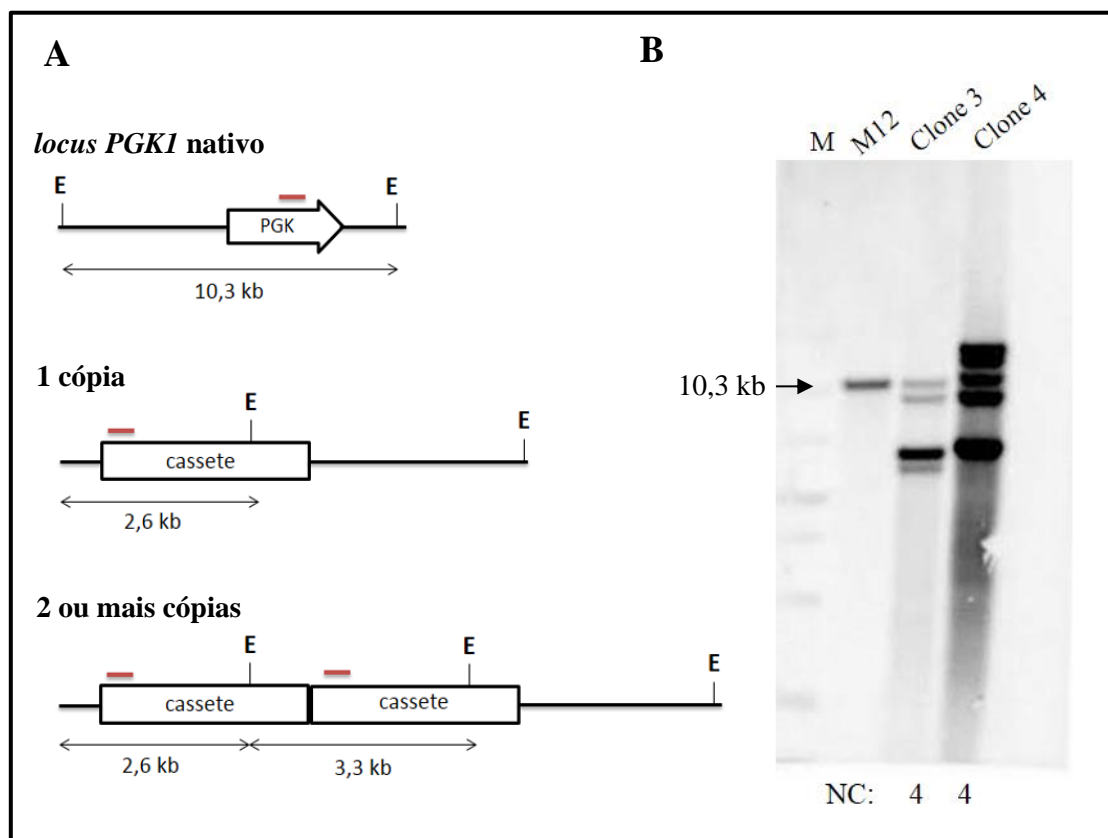
caso dos clones 3 e 4, o cassete foi integrado *in tandem* (21 e 20 cópias, respectivamente) em um único *locus* diferente ao *PGK1*, provavelmente, no cromossomo 2, como era esperado, porém, não foi observado o fragmento adicional de ~4,9 kb.

O DNA dos quatro clones analisados transformados com o cassete Chr3ld-GFP foi digerido com BglII. A sonda utilizada que hibridiza no promotor *PGK1* deveria permitir a identificação de um fragmento de ~11,8 kb se apenas uma cópia do vetor tivesse sido integrada. Se tivessem sido integradas duas cópias do vetor, seria identificado um outro fragmento de ~3,3 kb que corresponde ao tamanho do cassete e a intensidade desse fragmento iria aumentando por cada cópia a mais integrada. Um fragmento de ~6,2 kb correspondente ao *locus* do *PGK1* seria observado se a integração ocorresse em outro *locus* diferente. A figura 29A apresenta um esquema dos fragmentos esperados e a figura 29B mostra o resultado do Southern blot.



**Figura 29. Determinação do número de cópias dos clones transformados com o cassete Chr3ld-GFP.** A Esquema dos fragmentos esperados após digestão do gDNA com a enzima BglII, B Resultado do Southern blot. B: sítio de corte para a enzima BglII. NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). A linha vermelha indica o lugar onde hibridiza a sonda.

O fragmento de ~6,2 kb, que corresponde a uma cópia no genoma, apresentou uma intensidade maior nos clones 3 e 4, o que não permitiu determinar o número de cópias e, portanto, o DNA desses dois clones foi digerido com a enzima EcoRV. Com esta enzima, era esperado um fragmento de ~2,6 kb se apenas uma cópia do vetor tivesse sido integrada. Se tivessem sido integradas duas cópias do vetor seria identificado um outro fragmento de ~3,3 kb que corresponde ao tamanho do cassete e a intensidade desse fragmento iria aumentando por cada cópia a mais integrada. Um fragmento de ~10,3 kb corresponderia ao *locus* do *PGKI* intacto. A figura 30A apresenta um esquema dos fragmentos esperados e a figura 30B mostra o resultado do Southern blot.



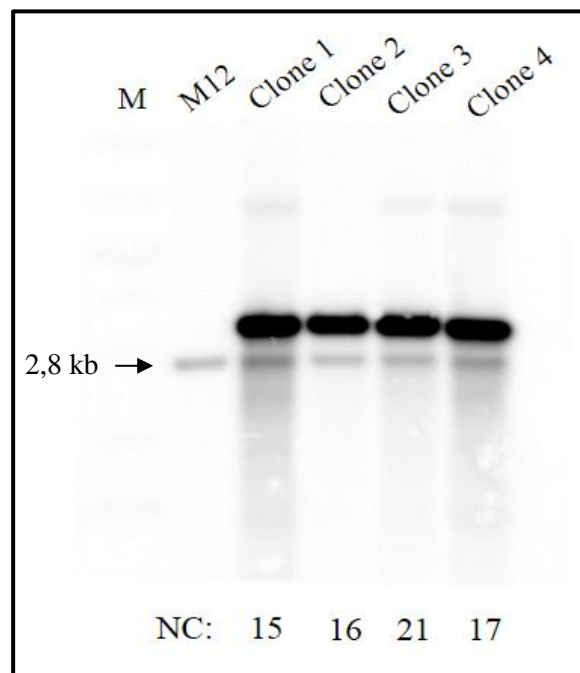
**Figura 30. Determinação do número de cópias dos clones 3 e 4 transformados com o cassete Chr3Id-GFP.** **A** Esquema dos fragmentos esperados após digestão do gDNA com a enzima EcoRV, **B** Resultado do Southern blot. E: sítio de corte para a enzima EcoRV. NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). A linha vermelha indica o lugar onde hibridiza a sonda.

A digestão do DNA dos clones 1 e 2 com BglII permitiu identificar o fragmento de ~6,2 kb, indicando que a integração não ocorreu no *locus PGKI*, porém, pode ter ocorrido em outro *locus* diferente do cromossomo 3 já que foram observados outros fragmentos além dos

esperados. A presença do fragmento de ~3,3 kb confirmou a integração de múltiplas cópias *in tandem* para ambos os clones. Usando como referência o fragmento de ~6,2 kb foi determinado o número de cópias: 72, para o clone 1 e 19, para o clone 2.

A digestão do DNA dos clones 3 e 4 com EcoRV também confirmou que a integração não ocorreu no *locus PGK1* já que foi observado o fragmento de ~10,3 kb correspondente ao *locus* intacto. O fragmento de ~3,3 kb correspondente ao cassete não foi identificado, enquanto que foi observado um fragmento de um pouco menos de 7 kb que poderia conter dois cassetes *in tandem* devido a digestão parcial. Outros fragmentos não esperados também foram observados e todos eles foram comparados com o fragmento de ~10,3 kb para determinar o número de cópias: 4 cópias para cada clone.

O DNA dos clones analisados transformados com o cassete 5SId-GFP foi digerido com a enzima EcoRI. Para a marcação foi usada a sonda que hibridiza no promotor *PGK1*. A figura 31 apresenta o resultado do Southern blot.

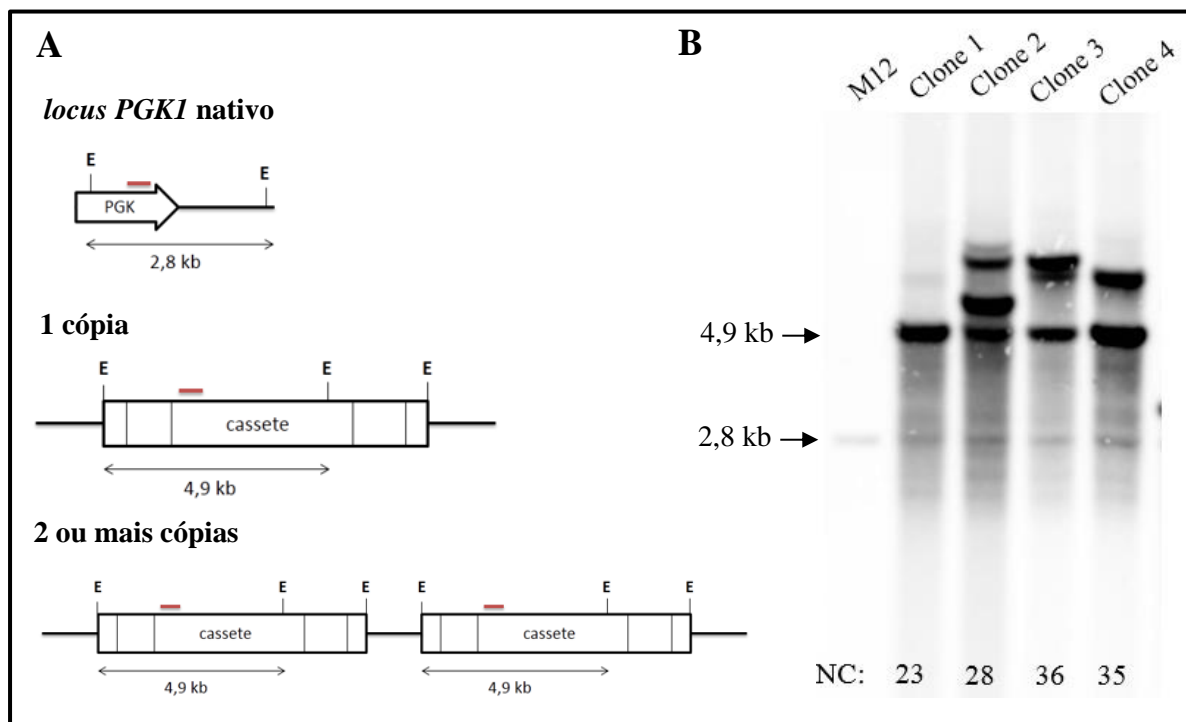


**Figura 31. Determinação do número de cópias dos clones transformados com o cassete 5SId-GFP.** Resultado do Southern blot. gDNA digerido com a enzima EcoRI. NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific).

Nesse caso, não foi possível determinar os tamanhos dos fragmentos que deveriam ser liberados já que o *locus* do rDNA 5S está espalhado no genoma, mas era esperado observar vários fragmentos de diferentes tamanhos. Porém, os quatro clones apresentaram um único

fragmento de ~3,3 kb, além do fragmento de ~2,8 kb correspondente ao *locus PGKI* intacto. Isto indicou que, apesar de existirem 21 repetições do rDNA 5S, todos os clones integraram várias cópias do cassete *in tandem* em um único *locus* e não houve grandes diferenças entre o número de cópias integradas em cada clone: 15, 16, 21 e 17 para os clones 1, 2, 3 e 4, respectivamente. Este resultado é compatível com a intensidade de fluorescência mais homogênea observada e com a curva de crescimento obtidas.

O DNA dos clones analisados transformados com o cassete NTSId-GFP foi digerido com a enzima EcoRI. A sonda utilizada que hibridiza no promotor *PGKI* devia permitir a identificação de um fragmento de ~4,9 kb se apenas uma cópia do vetor tivesse sido integrada. Se tivessem sido integradas múltiplas cópias, a intensidade desse fragmento de ~4,9 kb iria aumentando por cada cópia a mais integrada. Um fragmento de ~2,8 kb correspondente ao *locus* do *PGKI* seria observado se a integração ocorresse em outro *locus* diferente. A figura 32A apresenta um esquema dos fragmentos esperados e a figura 32B mostra o resultado do Southern blot.



**Figura 32. Determinação do número de cópias dos clones transformados com o cassete NTSId-GFP.** A Esquema dos fragmentos esperados após digestão do gDNA com a enzima EcoRI, B Resultado do Southern blot. E: sítio de corte para a enzima EcoRI. NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). A linha vermelha indica o lugar onde hibridiza a sonda.

Os quatro clones apresentaram o fragmento de ~2,8 kb correspondente ao *locus PGK1* intacto, indicando que não houve integração nesse *locus*. O fragmento de ~4,9 kb confirmou a integração múltipla *in tandem* no *locus* NTS do rDNA, porém, foram observados outros fragmentos não esperados que poderiam indicar integração de algumas cópias do cassete em outro *locus*. Usando como referência o fragmento de ~2,8 kb, foi determinado o número de cópias para cada clone: 23, 28, 36 e 35 cópias para os clones 1, 2, 3 e 4, respectivamente.

Diferentemente do observado com os clones transformados com o vetor pKld-GFP quando foram usados os cassetes com sequências repetidas, não foi observada uma correlação entre o número de cópias e a velocidade específica máxima de crescimento, nem entre os 16 clones, ou entre os quatro clones de cada sistema (figura 33) e, no caso dos clones transformados com o cassete Chr2ld-GFP, foi observado uma diminuição na taxa máxima de crescimento quando o número de cópias integradas foi maior.

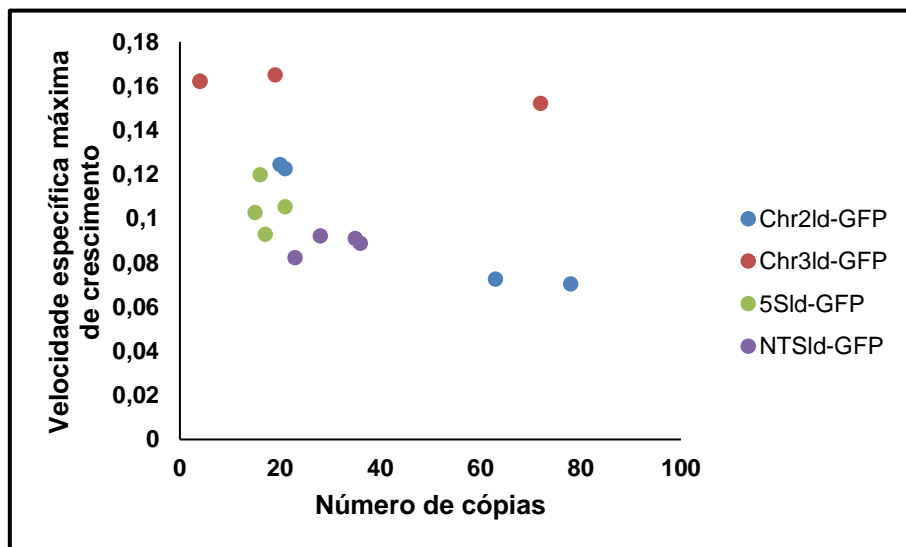
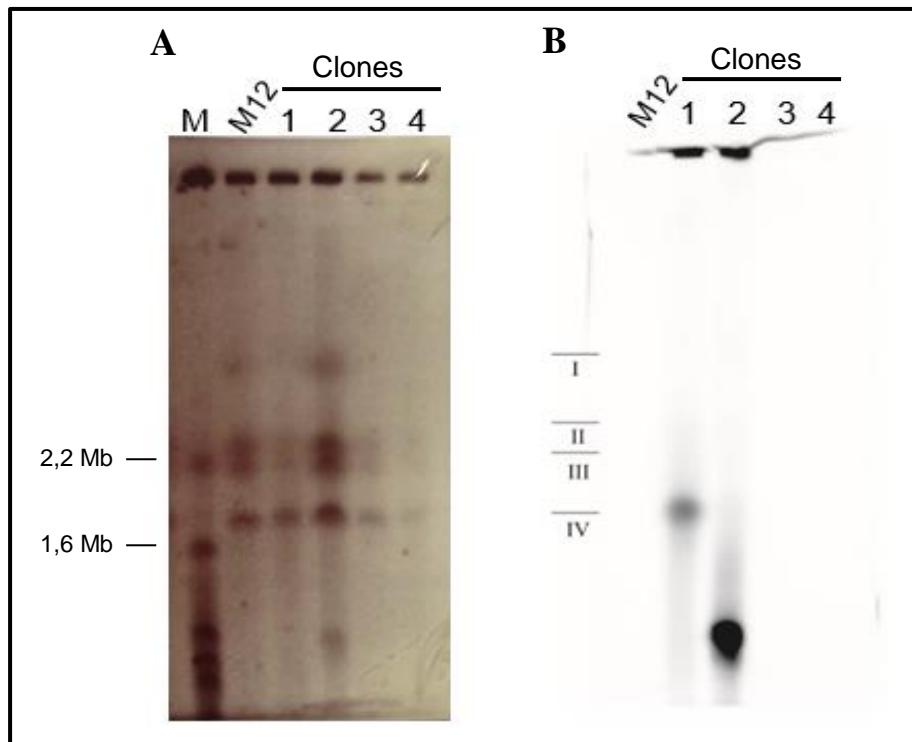


Figura 33. Correlação entre a velocidade específica máxima de crescimento e o número de cópias dos clones transformados com os cassetes contendo sequências repetidas e o gene EGFP.

As diferentes sequências repetidas presentes nos cassetes de expressão deviam direcionar a integração aos respectivos *loci* no genoma da levedura, porém, em todos os sistemas foi observado integração aleatória do cassete. Esse resultado também foi observado com o clone 5 (vetor pKld-GFP) que integrou em outro *locus* diferente apesar de ter sido direcionado ao *locus PGK1*. Contudo, essa integração aleatória não é inesperada já que em *K. phaffii*, além da recombinação homóloga, a integração pode estar mediada pelo sistema de

*non-homologous end joining* (NHEJ) o qual é o principal mecanismo de reparo em fungos filamentosos e eucariotos superiores (Näätsaari *et al.*, 2012). Além disso, a frequência com que acontecem eventos de substituição gênica por recombinação homóloga em *K. phaffii* e *K. pastoris*, depende do tamanho das sequências usadas como alvo de integração. Quando essas sequências têm menos de 500 pb a frequência de substituição gênica é de menos de 0,1% (Higgins e Cregg 1998). Excluindo o fragmento NTS rDNA, que possui aproximadamente 2500 pb, as sequências usadas neste trabalho apresentam tamanhos menores que 150 pb, portanto, era muito provável que a recombinação ocorresse em outros *loci* com certa frequência. Apesar de que a integração não foi 100% específica, a presença dessas sequências repetidas levou a um aumento do número de cópias do cassete de expressão que foram integradas quando comparado com a integração do vetor direcionado ao *locus PGK1*.

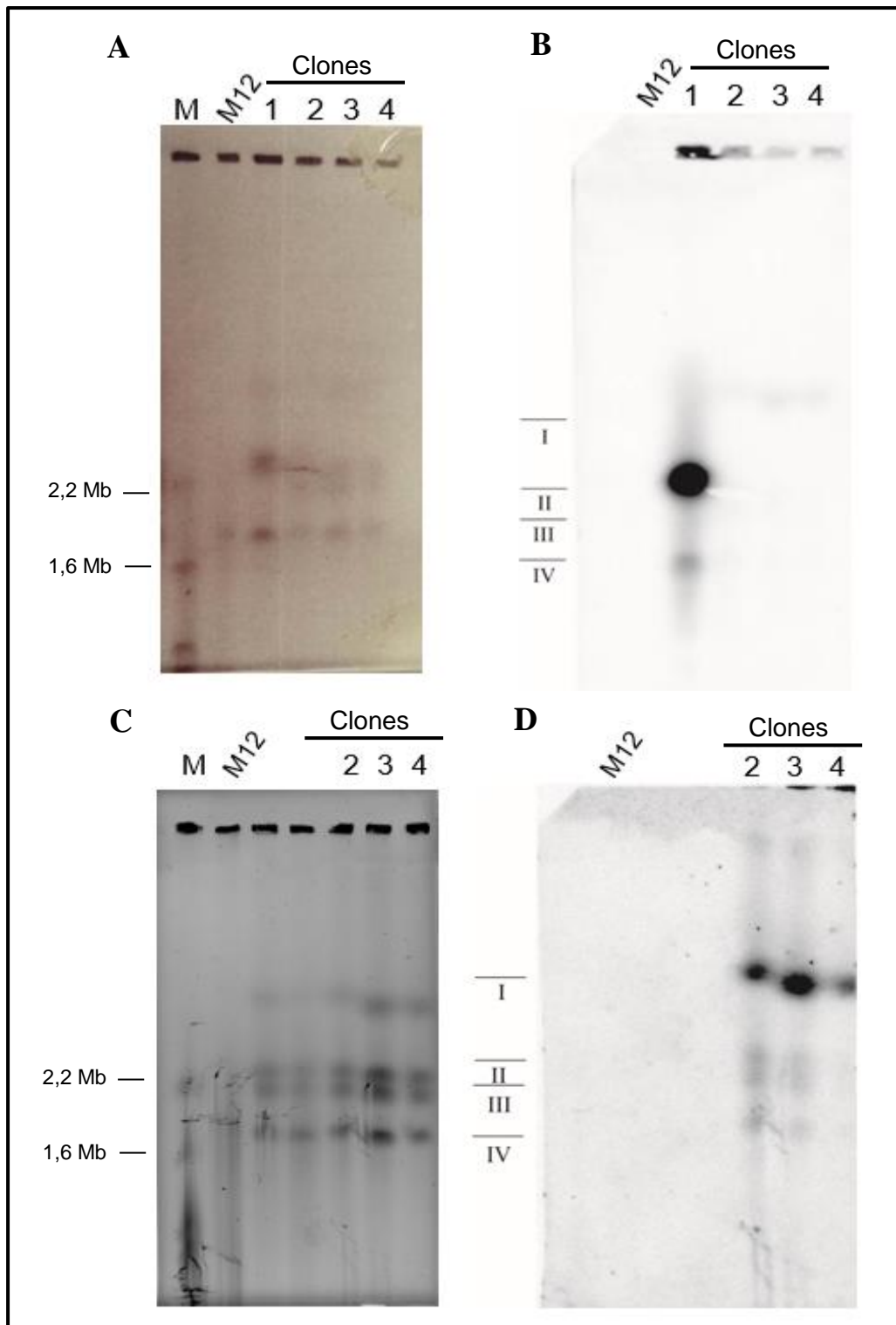
A fim de identificar em que lugar foram integradas as múltiplas cópias do cassete de expressão foi usada a técnica de eletroforese em campo pulsado para separar os cromossomos de cada um dos quatro clones selecionados para cada sistema. O DNA foi transferido a uma membrana de nitrocelulose e analisado por Southern blot usando uma sonda que hibridizava com o gene EGFP. O tamanho aproximado dos cromossomos 1, 2, 3 e 4 de *K. phaffii* é de 2,9 Mb; 2,4 Mb; 2,2 Mb e 1,8 Mb, respectivamente (Love *et al.*, 2016). A figura 34 apresenta o resultado da eletroforese (painel A) e o resultado do Southern blot (painel B) para os clones transformados com o cassete Chr2ld-GFP. Segundo observado na figura 34B, o clone 1 apresentou cópias do cassete de expressão integradas no cromossomo 4. O clone 2 apresentou um fragmento que não corresponde a nenhum dos quatro cromossomos de *K. phaffii* e esse fragmento extracromossomal é o que contém as cópias do cassete de expressão. Esse fragmento poderia ser um elemento de DNA circular (eccDNA) originado pela integração das múltiplas cópias do cassete perto a uma sequência ARS, como foi observado em *S. cerevisiae* (Demeke *et al.*, 2015). Quando o gene heterólogo *XylA* foi integrado perto de uma sequência ARS no genoma de *S. cerevisiae*, foi formado um eccDNA seguido de uma nova integração no genoma, aumentando o número de cópias *in tandem* durante o processo de evolução adaptativa. O eccDNA contendo o gene *XylA* foi extraído da levedura e transferido para outra linhagem de *S. cerevisiae* a qual foi capaz de fermentar xilose, confirmando a estabilidade do fragmento extracromossomal. No nosso trabalho, um elemento de DNA circular estável poderia ter sido originado no clone 2 se as 63 cópias do cassete tivessem sido integradas perto de uma sequência ARS de *K. phaffii*.



**Figura 34. Determinação do local de integração do cassete Chr2ld-GFP.** **A** Resultado da eletroforese em campo pulsado em gel de agarose 1% corado com brometo de etídeo. M: marcador (cromossomos de *S. cerevisiae*), **B** Resultado do Southern blot. Os números I, II, III e IV indicam as posições dos cromossomos 1, 2, 3 e 4 respectivamente.

Especulamos que nos clones 3 e 4 não foram identificados fragmentos devido ao menor número de cópias integradas nesses clones (21 e 20 cópias, respectivamente) o que poderia ter gerado um sinal muito fraco ao contrário dos clones 1 e 2 com um alto número de cópias (78 e 63 cópias, respectivamente), porém, quando estes foram analisados separadamente em outro gel, também não houve nenhum sinal de detecção. Desta forma, para os clones 3 e 4, não foi possível determinar o local de integração. Todos esses resultados mostraram que, de fato, o cassete de expressão contendo sequências repetidas do cromossomo 2 não foi integrado no *locus* para o qual foi direcionado.

A figura 35 apresenta o resultado da eletroforese (painéis A e C) e o resultado do Southern blot (painéis B e D) para os clones transformados com o cassete Chr3ld-GFP. Os painéis A e B apresentam os resultados para os quatro clones analisados, porém, apenas o clone 1 apresentou fragmentos. Como os clones 2, 3 e 4 apresentam menor número de cópias, especulamos que não foram observados fragmentos pois o sinal foi muito fraco em comparação com o clone 1 e, portanto, esses três clones foram analisados separadamente em outro gel (painéis C e D).

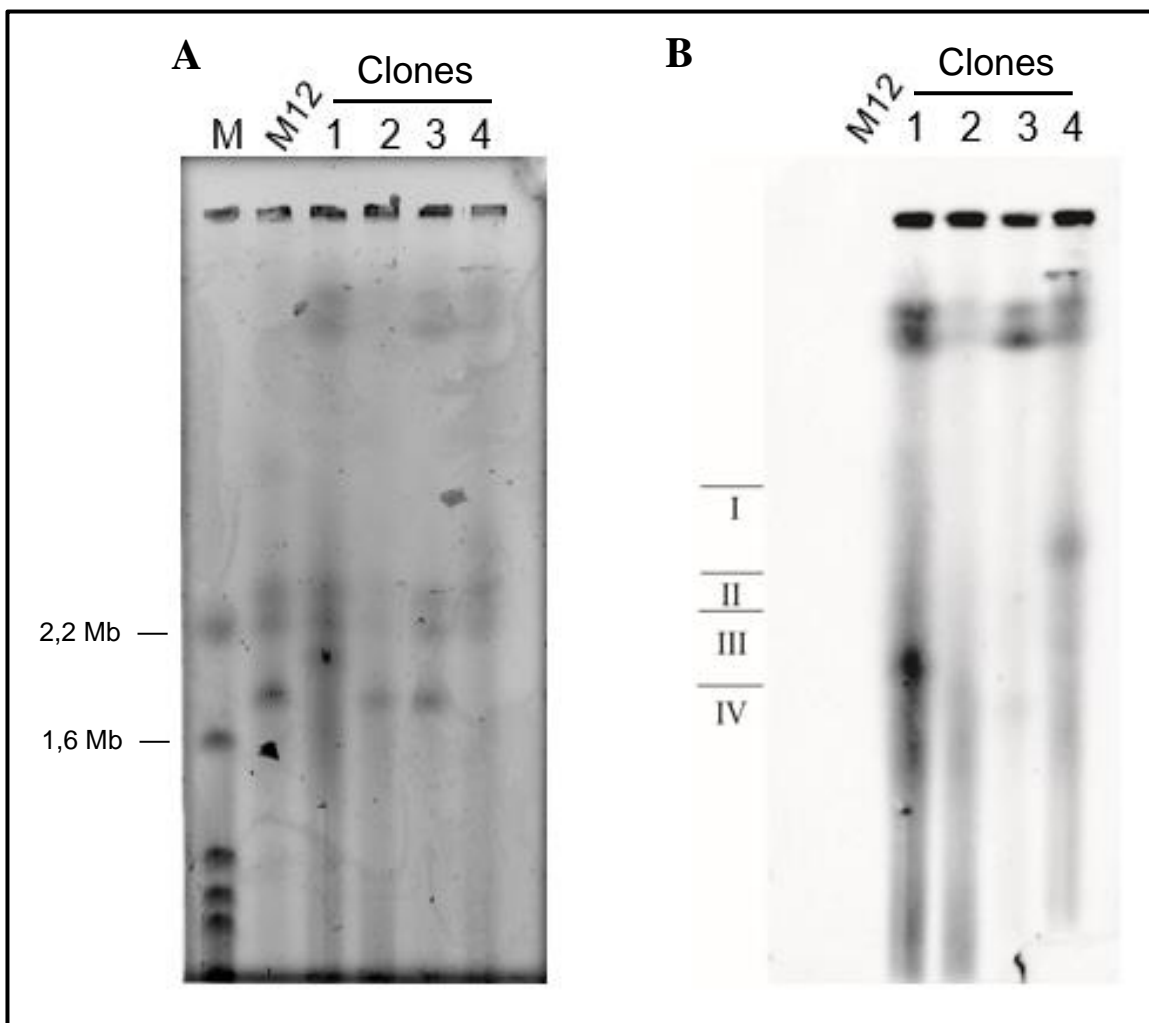


**Figura 35. Determinação do local de integração do cassete Chr3ld-GFP.** **A** Resultado da eletroforese em campo pulsado em gel de agarose 1% corado com brometo de etídeo dos quatro clones selecionados, **B** Resultado do Southern blot para os quatro clones selecionados, **C** Resultado da eletroforese em campo pulsado em gel de agarose 1% corado com brometo de etídeo dos clones 2, 3 e 4, **D** Resultado do Southern blot para os clones 2, 3 e 4. Os números I, II, III e IV indicam as posições dos cromossomos 1, 2, 3 e 4, respectivamente. M: marcador (cromossomos de *S. cerevisiae*).



Segundo observado na figura 35B, o clone 1 apresentou a maioria das cópias do cassete de expressão integradas no cromossomo 2 e umas poucas cópias no cromossomo 4. O resultado da figura 35D mostra que os clones 2, 3 e 4 apresentaram cópias do cassete integradas no cromossomo 1. Estes resultados indicam que o cassete de expressão contendo sequências repetidas do cromossomo 3 também não foi integrado no *locus* para o qual foi direcionado.

A figura 36 apresenta o resultado da eletroforese (painel A) e o resultado do Southern blot (painel B) para os clones transformados com o cassete NTSId-GFP.



**Figura 36. Determinação do local de integração do cassete NTSId-GFP.** **A** Resultado da eletroforese em campo pulsado em gel de agarose 1% corado com brometo de etídeo. M: marcador (cromossomos de *S. cerevisiae*), **B** Resultado do Southern blot. Os números I, II, III e IV indicam as posições dos cromossomos 1, 2, 3 e 4 respectivamente.

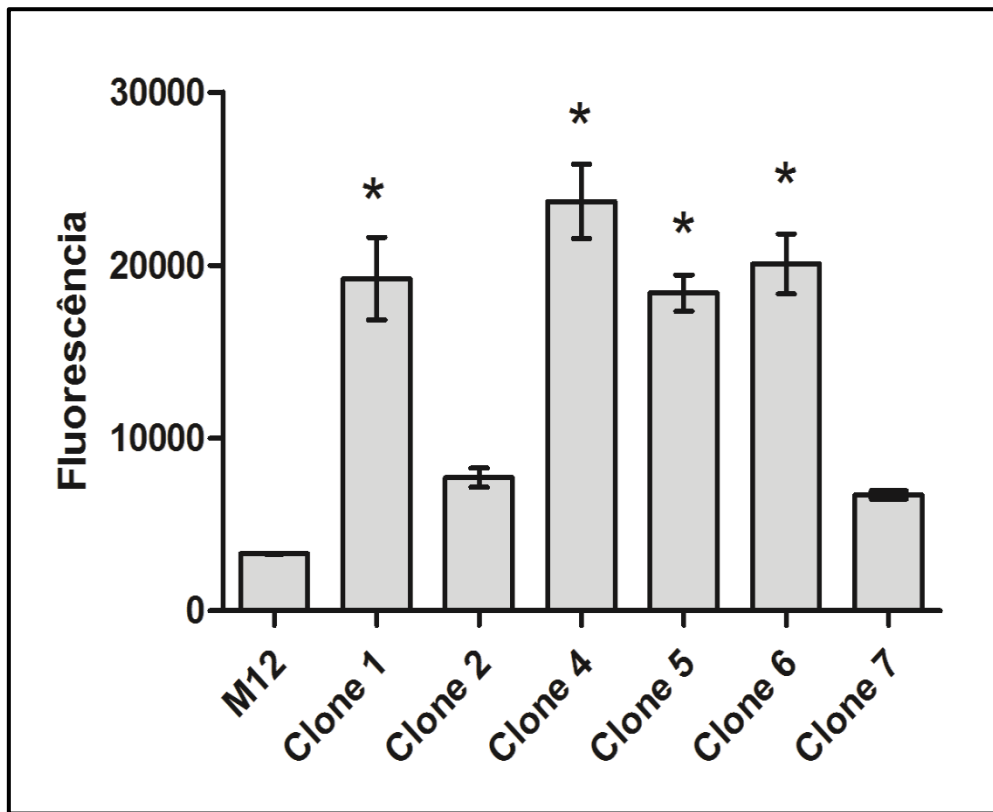
Como observado na figura 36A, a banda correspondente ao cromossomo 1, presente no clone M12, não foi observada nos clones transformantes. Dois fragmentos com tamanhos muito maiores foram observados para os quatro clones e, como mostrado na figura 36B, esses fragmentos são os que contêm as múltiplas cópias do cassete de expressão. Recentemente, foi evidenciado que o DNA ribossomal da linhagem *K. phaffii* GS115 encontra-se unicamente no cromossomo 1 (Love *et al.*, 2016) e, portanto, podemos especular que os fragmentos observados no Southern blot correspondem ao cromossomo 1 que aumentou de tamanho nos quatro clones analisados. Em *S. cerevisiae* é conhecido que o rDNA pode sofrer expansões e contrações durante o processo de replicação (Kobayashi *et al.*, 1998; Kobayashi *et al.*, 2001). Apesar de o mecanismo de replicação do rDNA ainda não estar descrito em *K. phaffii*, já foi demonstrado que a integração nesse *locus* leva a uma amplificação do vetor de expressão sob pressão seletiva (Marx *et al.*, 2009). É possível que *K. phaffii* tenha um mecanismo similar ao de *S. cerevisiae* e que na região NTS existem elementos reconhecidos por enzimas encarregadas de manter estável o número de repetições. A integração no NTS pode ter interrompido um desses elementos e isso poderia explicar o incremento do tamanho dessa região no cromossomo 1.

A localização do cassete de expressão nos clones transformados com o cassete 5SId-GFP não pôde ser determinada, apesar de haver identificado os cromossomos no gel de agarose, não foi observado nenhum fragmento na membrana de nitrocelulose nas três vezes que foi repetido o experimento.

### 5.2.5 Produção de GFP

Os seis clones transformados com o vetor pKId-GFP foram analisados por citometria de fluxo para quantificar a produção de GFP. A figura 37 apresenta a mediana de fluorescência das células positivas para GFP (células produzindo a proteína). Cada coluna corresponde à média de três experimentos para cada clone e a barra de erro representa o erro padrão da média. Todos os clones transformantes apresentaram valores de fluorescência maiores ao da levedura não-transformada (M12) indicando que a presença do gene EGFP é responsável pela fluorescência observada, porém, apenas os clones 1, 4, 5 e 6 mostraram diferença significativa. Os clones 2 e 7, que têm um crescimento mais lento e um menor número de cópias, mostraram os menores níveis de produção da proteína heteróloga, sendo

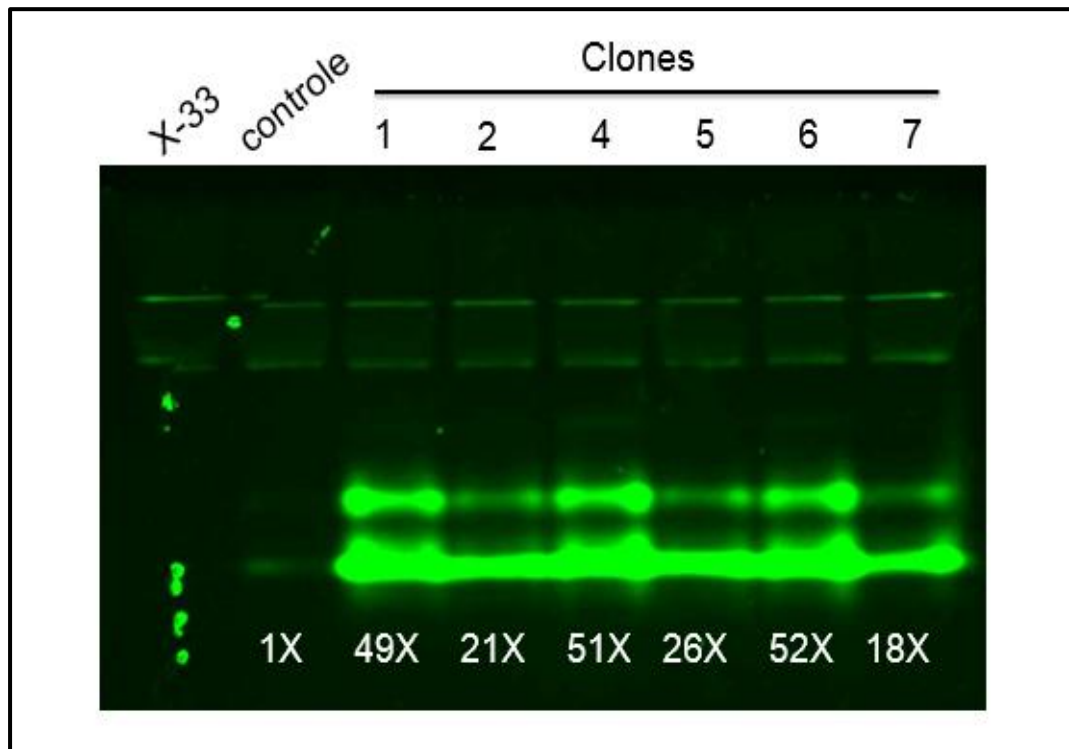
que, estatisticamente, estes valores não apresentaram diferença significativa comparados com a fluorescência apresentada pela linhagem M12. Contudo, é importante notar que a porcentagem de células de M12 positivas para GFP foi menor que 1%.



**Figura 37. Produção intracelular de GFP em clones transformados com pKld-GFP.** Análise da produção por citometria de fluxo. O valor apresentado corresponde à média de fluorescência para  $n=3$ . As barras de erro mostram o erro padrão da média. Os asteriscos indicam os clones que apresentaram diferença significativa em comparação a *K. phaffii* M12 ( $p < 0,05$ ).

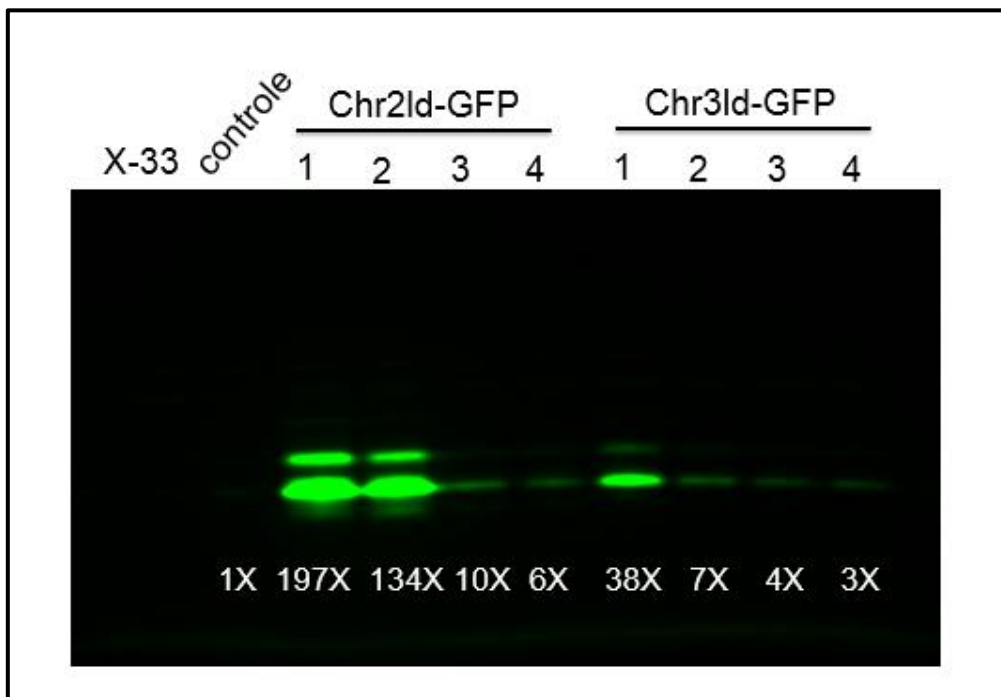
A fluorescência observada em cada clone também foi comparada com aquela de um clone controle contendo uma cópia do gene EGFP para obter um valor relativo de produção intracelular de GFP em gel não-desnaturante. Duas bandas foram observadas no gel devido à tendência de GFP a formar dímeros (Costantini *et al.*, 2013) e ambas as bandas foram usadas para determinar os valores de produção relativa (figura 38). Os clones 1, 4 e 6 foram os que apresentaram maiores valores de produção relativa. Este resultado está em concordância com os valores de fluorescência obtidos por citometria de fluxo, porém, o clone 5 mostrou uma produção relativa baixa, similar à apresentada pelos clones 2 e 7, que tiveram menores valores

de fluorescência na citometria de fluxo. Contudo, cabe notar que o experimento do gel não-desnaturante foi feito uma única vez, enquanto que na citometria de fluxo foram feitas três réplicas do experimento, o que daria valores mais confiáveis. É possível que durante a lise celular pudesse ter havido alguma perda ou degradação da proteína do clone 5, o que poderia explicar a diminuição da fluorescência observada no gel.

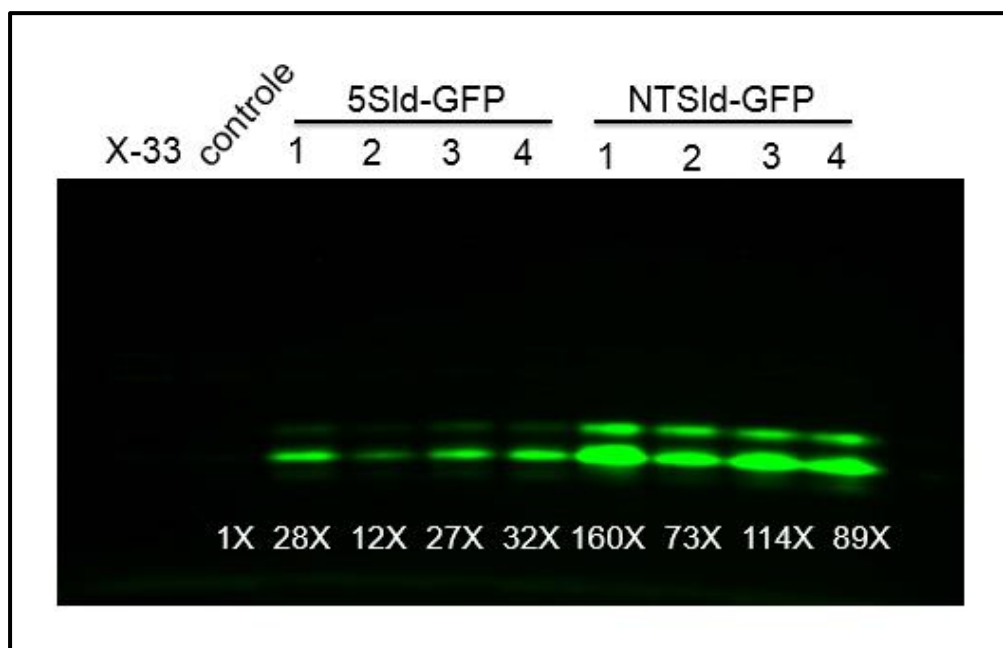


**Figura 38. Produção intracelular de GFP em clones transformados com o vetor pKld-GFP.** Análise da fluorescência em gel não desnaturante. Produção relativa em comparação ao clone controle (clone de *K. phaffii* contendo uma cópia do gene EGFP).

Com relação aos clones transformados com os cassetes Chr2ld-GFP, Chr3ld-GFP, 5Sld-GFP e NTSld-GFP, a produção relativa de GFP foi analisada em gel não-desnaturante. Como observado nas figuras 39 e 40, todos os clones avaliados produziram mais GFP do que o clone contendo uma única cópia. Da mesma forma como observado com os clones transformados com o vetor pKld-GFP, a maior produção de proteína foi identificada nos clones com maior número de cópias, sendo que os clones 1 e 2 (transformados com o cassete Chr2ld-GFP) foram os que apresentaram maior produção de GFP. A tabela 9 resume os valores de produção relativa de GFP assim como o número de cópias integradas do vetor ou do cassete e a velocidade específica máxima de crescimento para cada clone analisado



**Figura 39. Produção intracelular de GFP em clones transformados com os cassetes Chr2Id-GFP e Chr3Id-GFP.** Análise da fluorescência em gel não desnaturante. Produção relativa em comparação ao clone controle (clone de *K. phaffii* contendo uma cópia do gene EGFP).



**Figura 40. Produção intracelular de GFP em clones transformados com os cassetes 5SId-GFP e NTSId-GFP.** Análise da fluorescência em gel não desnaturante. Produção relativa em comparação ao clone controle (clone de *K. phaffii* contendo uma cópia do gene EGFP).

**Tabela 9.** Relação de número de cópias, produção relativa de GFP e velocidade de crescimento

Sistema	Clone	Número de cópias	Produção relativa de GFP*	$\mu_{\text{máx}}$
	controle	1	1	
pKld-GFP	1	17	49	0,1596
	2	7	21	0,0708
	4	20	51	0,1614
	5	14	26	0,1408
	6	17	52	0,1650
	7	5	18	0,0502
	Chr2ld-GFP	1	78	197
2		63	134	0,0725
3		21	10	0,1225
4		20	6	0,1244
Chr3ld-GFP	1	72	38	0,1521
	2	19	7	0,1650
	3	4	4	0,1620
	4	4	3	0,1621
5Sld-GFP	1	15	28	0,1026
	2	16	12	0,1198
	3	21	27	0,1052
	4	17	32	0,0928
NTSld-GFP	1	23	160	0,0822
	2	28	73	0,0920
	3	36	114	0,0888
	4	35	89	0,0909

\* Produção em relação ao clone com uma cópia do gene.

Apesar de o clone 1 (sistema Chr3ld-GFP) apresentar um número de cópias integradas do cassete similar ao integrado nos clones 1 e 2 (sistema Chr2ld-GFP), a produção de GFP

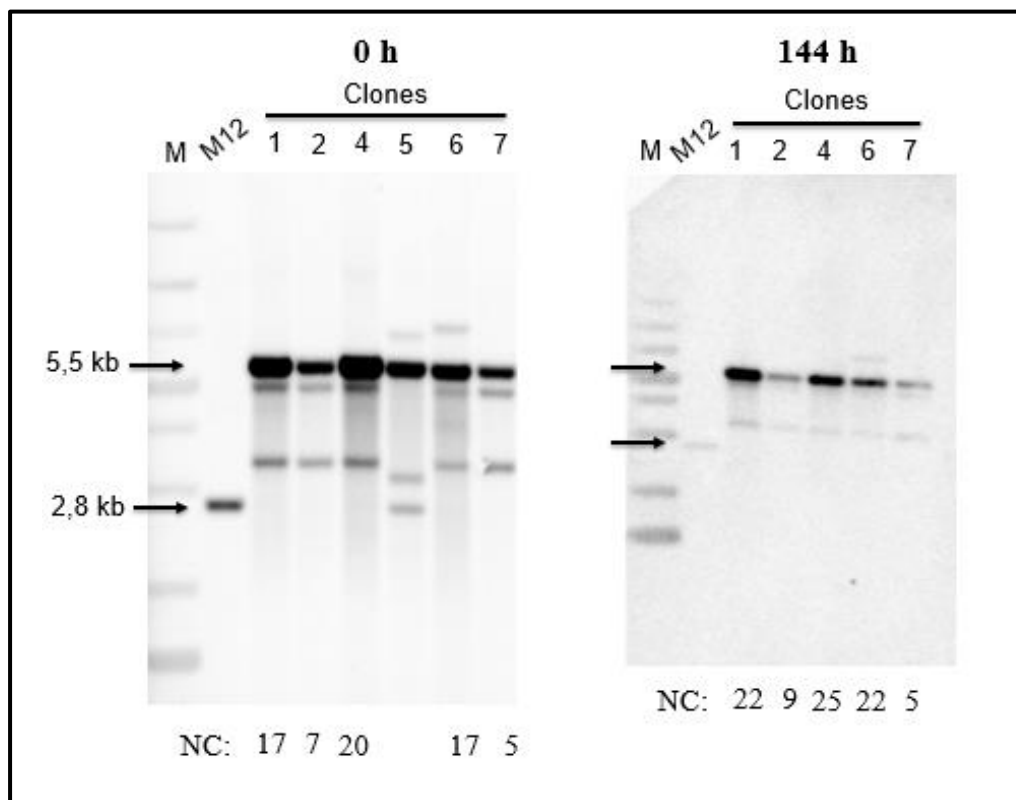
não foi tão alta quanto à produção nesses dois últimos clones. Isto pode ser explicado pelo fato de que a integração em cada clone foi em um local diferente. Desta forma, é provável que o cassete Chr3ld-GFP tenha sido integrado no clone 1 em um *locus* com uma estrutura genômica inibidora de transcrição. Esses baixos níveis de produção da proteína nos quatro clones do sistema Chr3ld-GFP também explicariam por que a velocidade específica máxima de crescimento não foi alterada, enquanto que os clones dos demais sistemas, com maiores níveis de produção apresentaram uma taxa de crescimento menor. Os clones 1 e 2 do sistema Chr2ld-GFP e os quatro clones do sistema NTSld-GFP foram os que mostraram menores taxas de crescimento e, ao mesmo tempo, foram os que apresentaram maiores níveis de produção de GFP. Esta correlação negativa entre número de cópias e velocidade de crescimento já foi observada em outros trabalhos (Cos *et al.*, 2005; Zhu *et al.*, 2009) onde a integração de múltiplas cópias do vetor de expressão em *K. phaffii* resultou em um efeito negativo no crescimento celular devido provavelmente a uma sobrecarga do metabolismo ou à toxicidade da proteína heteróloga produzida. Os resultados deste trabalho mostram que, ao usar o sistema aqui apresentado, deve-se buscar por colônias de maior tamanho assim como pelas colônias de tamanho médio, que são as que possuem um número ótimo de cópias integradas. Dependendo do local de integração, as colônias de maior tamanho (e maior velocidade de crescimento) podem ser as que apresentam um maior número de cópias integradas, como foi o caso dos clones transformados com o vetor pKld-GFP. Em outros casos, o alto número de cópias pode afetar o crescimento e, portanto, as colônias de tamanho médio (e uma velocidade de crescimento um pouco menor) são as desejadas. As colônias pequenas não devem ser selecionadas já que são as que possuem um baixo número de cópias integradas e, portanto, não recuperaram completamente a prototrofia, ou podem ser as que possuem um número elevado de cópias integradas, que pode ser prejudicial para a célula ao exigir muito mais da maquinaria celular.

Em resumo, a produção intracelular de proteína verde fluorescente foi confirmada nas células transformadas com o vetor pKld-GFP e com os cassetes Chr2ld-GFP, Chr3ld-GFP, 5Sld-GFP e NTSld-GFP, demonstrando que *K. phaffii* M12 pode ser transformada satisfatoriamente com um vetor ou cassete de expressão contendo o alelo *leu2-d* como marca de seleção. Além disso, a transformação com os cassetes de expressão confirmou que a presença do alelo *leu2-d* e das sequências repetidas no vetor permite a integração de múltiplas cópias do vetor no genoma da levedura.

### 5.2.6 Estabilidade genética

Os clones selecionados de cada sistema foram crescidos em meio não seletivo (YPD) durante 72 gerações para avaliar a estabilidade genética. A cultura foi transferida para meio fresco a cada 24 h para garantir que as células estivessem na fase exponencial durante todo o experimento. Após 144 h de crescimento, o número de cópias dos clones transformados com o vetor pKld-GFP foi analisado por Southern blot após digestão do DNA genômico com a enzima EcoRI. Como o clone 5 apresentou diferentes cópias do vetor integradas em outro *locus* diferente do *PGK1*, consideramos que não era passível de comparação com os outros clones desse sistema e, portanto, sua estabilidade genética não foi analisada.

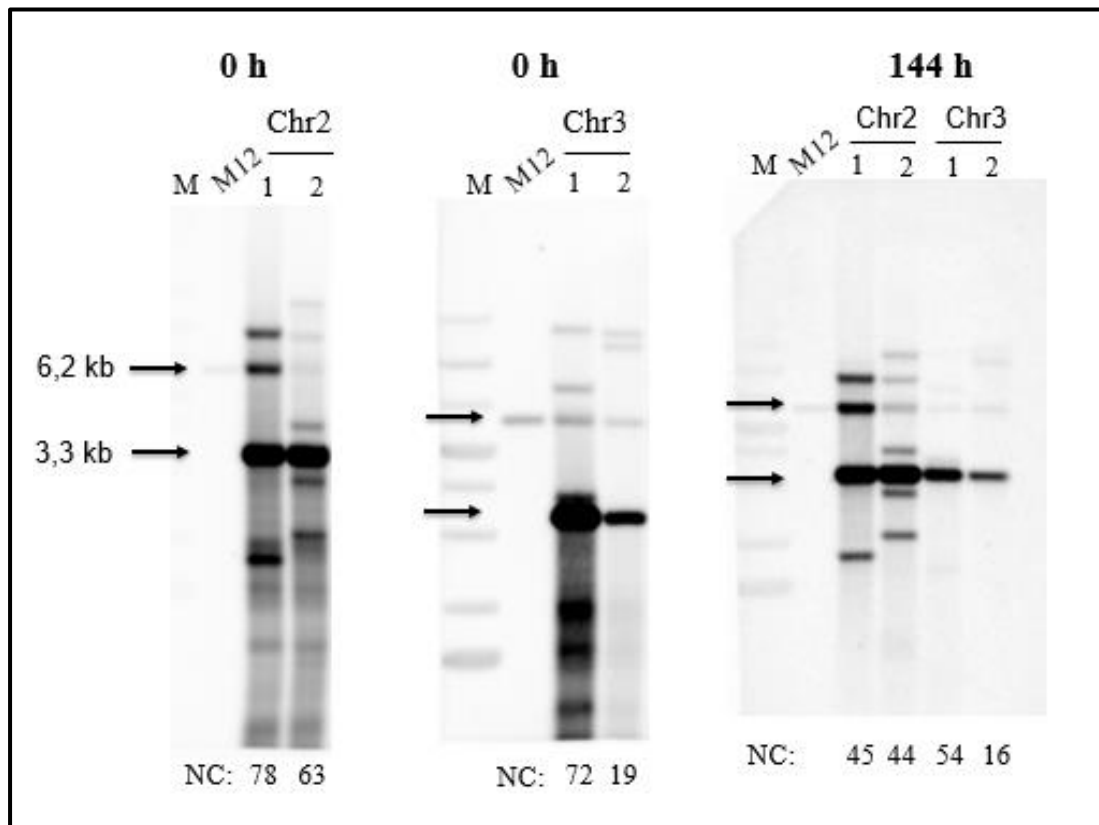
Como observado na figura 41, em nenhum dos clones avaliados houve diminuição do número de cópias do vetor, indicando que o vetor foi mantido estavelmente durante 72 gerações.



**Figura 41. Estabilidade genética dos clones transformados com o vetor pKld-GFP.** Resultado do Southern blot antes (0 h) e após 144 h de crescimento em meio não seletivo (YPD). NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). As setas indicam o fragmento de 5,5 kb correspondente ao vetor e o fragmento de 2,8 kb correspondente ao *locus* nativo de gene *PGK1* após digestão com EcoRI.



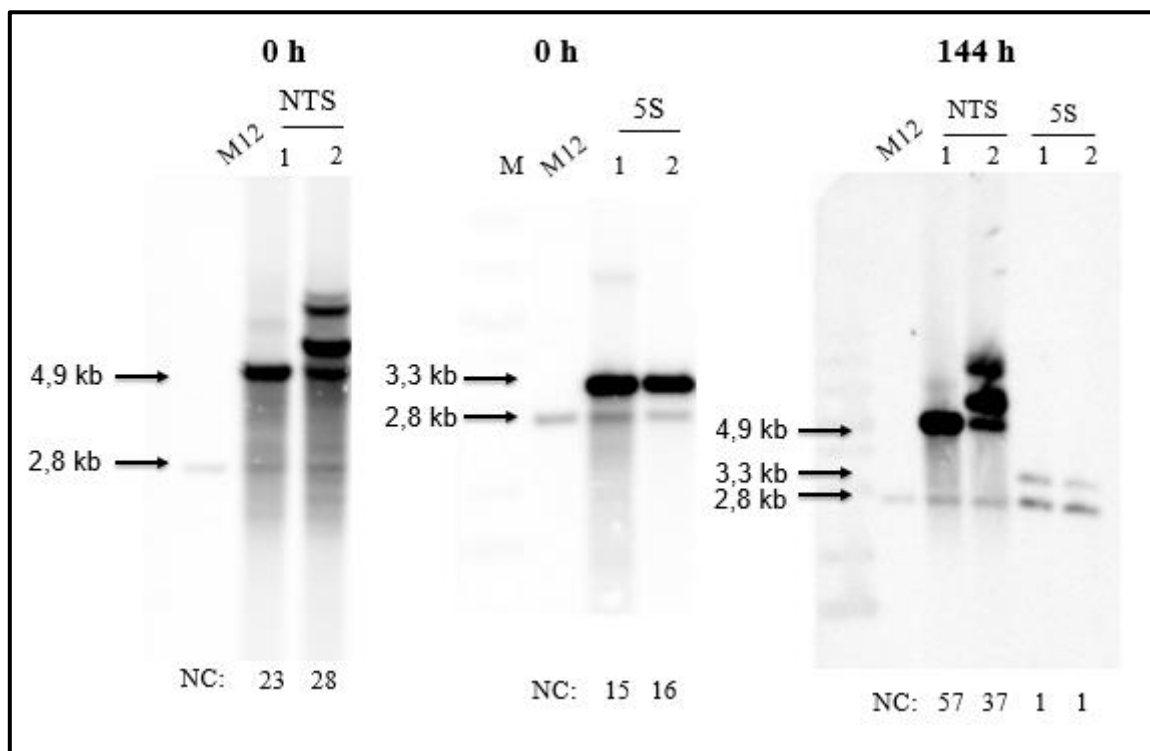
Após 144 h de crescimento, o número de cópias dos clones transformados com os diferentes cassetes foi analisado por Southern blot após digestão do DNA genômico com BglII, para os sistemas Chr2ld-GFP e Chr3ld-GFP (figura 42), e EcoRI, para os sistemas 5Sld-GFP e NTSld-GFP (figura 43). Para cada sistema foram escolhidos os dois clones com maior número de cópias para esta análise. Como pode ser observado na figura 42, tanto os clones transformados com o cassete Chr2ld-GFP como os transformados com Chr3ld-GFP houve perda de cópias do cassete, sendo que esta foi maior nos clones com alto número de cópias (mais de 60 cópias iniciais): os clones com 78, 63 e 72 cópias ficaram com 45, 44 e 54 cópias, o que corresponde a uma perda de 42%, 30% e 25%, respectivamente. O clone com 19 cópias iniciais ficou com 16, correspondente a uma perda de 16%.



**Figura 42. Estabilidade genética dos clones transformados com os cassetes Chr2ld-GFP e Chr3ld-GFP.** Resultado do Southern blot antes (0 h) e após 144 h de crescimento em meio não seletivo (YPD). NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). As setas indicam o fragmento de 3,3 kb correspondente ao cassete e o fragmento de 6,2 kb correspondente ao *locus* nativo de gene *PGK1* após digestão com BglII.

Como mostrado na figura 43, os clones transformados com o cassete NTSId-GFP aumentaram o número de cópias integradas do cassete, com um incremento de 148% e 32% para os clones com 23 e 28 cópias, respectivamente. Amplificações gênicas espontâneas já foram observadas em *S. cerevisiae*, como a dos genes *ADH2* e *ADH4* (Paquin *et al.*, 1992; Dorsey *et al.*, 1992).

No caso dos clones transformados com o cassete 5SId-GFP, os dois clones avaliados perderam quase todas as cópias do cassete restando uma única cópia após 72 gerações.



**Figura 43. Estabilidade genética dos clones transformados com os cassetes NTSId-GFP e 5SId-GFP.** Resultado do Southern blot antes (0 h) e após 144 h de crescimento em meio não seletivo (YPD). NC: número de cópias. M: marcador O'GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific). As setas indicam os fragmentos de 3,3 kb ou 4,9 kb correspondentes ao cassete e o fragmento de 2,8 kb correspondente ao *locus* nativo de gene *PGK1* após digestão com EcoRI.

Trabalhos anteriores mostraram que 1 cópia ou baixos números de cópias de vetores ou cassetes de expressão são geneticamente estáveis em *K. phaffii* sob diferentes condições. Por exemplo, duas cópias do gene HSA foram mantidas estávelmente no genoma da levedura durante 83 gerações (Ohi *et al.*, 1998), assim como duas cópias do gene que codifica a

proteína guamerina durante 70 gerações (Lim *et al.*, 2000). Contudo, existem poucos estudos focados na estabilidade de clones multicópia. Em um estudo recente, linhagens de *S. cerevisiae* com múltiplas cópias integradas de cassetes contendo diferentes marcas auxotróficas defectivas mostraram estabilidade mitótica sob condições não seletivas (Moon *et al.*, 2016). Contrariamente, células de *S. cerevisiae* transformadas com cinco ou mais cópias de um vetor integrativo para expressão do gene *SUC2* e contendo a marca que confere resistência a G418 mostraram ser muito instáveis em meio não-seletivo (Wang *et al.*, 1996). Quando *K. phaffii* foi transformada com uma série de vetores contendo diferentes números de cópias do gene precursor da insulina porcina (PIP), tanto os clones com baixo como com alto número de cópias foram estáveis em meio não seletivo (YPD). Contudo, os clones com alto número de cópias perderam cópias do cassete contendo o gene PIP depois de 96 h de indução em meio com metanol (Zhu *et al.*, 2009).

A estabilidade dos arranjos *in tandem* pode estar afetada pela recombinação intramolecular que pode eliminar o gene de interesse na falta de pressão seletiva (Lee e Da Silva 1997). Desta forma, era esperado que todos os clones obtidos neste trabalho apresentassem pouca estabilidade já que a integração ocorreu *in tandem* para todos os sistemas, inclusive quando usada a sequência 5S do rDNA com a qual era esperada uma integração ectópica. No entanto, os nossos resultados mostraram que quando a integração ocorreu no *locus PGK1*, clones contendo até 20 cópias do vetor foram estáveis, porém, quando a integração foi dirigida as sequências repetidas, houve perda de algumas cópias, sendo que o *locus* 5S do rDNA mostrou ser uma região pouco estável.

A expressão intracelular do gene EGFP em *K. phaffii* M12 mostrou a viabilidade de se usar esta linhagem junto com vetores contendo a marca *leu2-d* como uma plataforma para a produção de proteínas heterólogas e demonstrou também a possibilidade de se usar as sequências repetidas do cromossomo 2, do cromossomo 3 e as sequências 5S e NTS do rDNA como alvos de integração para obter linhagens multicópia. A fim de verificar o perfil desses vetores em expressão extracelular foi escolhida o gene *AMY1* para essa análise

### 5.3 Integração múltipla do gene *AMY1*

A  $\alpha$ -amilase encontra-se entre as mais importantes enzimas hidrolíticas com aplicação em diferentes indústrias. Esta enzima, que cliva aleatoriamente as ligações glicosídicas  $\alpha$ -1,4 presentes na parte interna do amido, é usada como aditivo na fabricação de pão para aumentar o volume, a textura e o sabor do produto. Além disso, os hidrolisados de amido são utilizados na indústria alimentícia como adoçantes, espessantes e para a obtenção de xaropes empregados em confeitarias, cervejarias, fabricação de geleias e de refrigerantes. Adicionalmente as  $\alpha$ -amilases são usadas como aditivo em detergentes, na modificação do amido para fabricação de papel revestido e na indústria têxtil para remover o amido que é usado como camada protetora dos tecidos. Uma outra aplicação das  $\alpha$ -amilases é na produção de etanol como biocombustível. O amido é o principal substrato para esse processo, porém, ele deve ser sacarificado usando  $\alpha$ -amilases para que microrganismos fermentadores como *S. cerevisiae* utilizem os açúcares fermentáveis para produzir etanol (Gupta *et al.* 2003; Souza e Magalhães 2010).

Na busca por micro-organismos amilolíticos na biodiversidade do Brasil foi isolada a levedura *Cryptococcus flavus* e uma amilase secretada por este organismo foi caracterizada bioquimicamente (Wanderley *et al.*, 2004). Esta enzima, chamada de Amy1, apresenta propriedades importantes para aplicações biotecnológicas: baixo *K<sub>m</sub>* para amido solúvel (0,0056 mg/mL), alta estabilidade a pH 5,5 e temperatura ótima de 55°C. Num trabalho posterior, o gene *AMY1* de *C. flavus* foi clonado e expresso satisfatoriamente em *S. cerevisiae* (Galdino *et al.*, 2008). A sequência de nucleotídeos do cDNA revelou uma ORF de 1896 pb que codifica um polipeptídeo de 631 resíduos de aminoácidos o qual apresentou alta identidade quando comparado com a sequência de aminoácidos da amilase amy-CS2 isolada de *Cryptococcus* sp. S-2 (Iefuji *et al.*, 1996). A proteína recombinante tem uma massa molecular aparente de 67 kDa, similar à massa molecular da enzima nativa. A caracterização bioquímica e estrutural da proteína recombinante produzida em *S. cerevisiae* mostrou algumas diferenças significativas em comparação com a proteína nativa, principalmente diferentes padrões de dobramento o que viu-se refletido na diminuição da atividade da proteína recombinante e no aumento do *K<sub>m</sub>* (0,37 mg/mL) usando amido solúvel como substrato (Galdino *et al.*, 2011). Apesar dessas diferenças, a  $\alpha$ -amilase recombinante apresenta potencial biotecnológico devido a sua alta termoestabilidade e a sua alta atividade nos substratos amido, amilopectina e amilose, enquanto que a proteína nativa mostrou baixa

atividade nos dois últimos. No presente trabalho, o gene *AMY1* de *C. flavus* foi clonado nos vetores contendo as diferentes sequências repetidas, como modelo para estudar a produção de uma proteína secretada usando as estratégias aqui apresentadas.

### 5.3.1 Construção dos vetores pBSChr2ld-AMY, pBSChr3ld-AMY, pBS5Sld-AMY e pPCVNTSld-AMY

A estratégia usada para construir os vetores com sequências repetidas contendo o gene *AMY1* de *C. flavus* está apresentada na figura 44. Os vetores pBSChr2ld-GFP, pBSChr3ld-GFP, pBS5Sld-GFP e pPCVNTSld-GFP foram digeridos com BamHI e NotI para remover o gene EGFP e, no lugar, foi clonado o gene *AMY1* fusionado ao peptídeo sinal do fator  $\alpha$  de *S. cerevisiae* com códons otimizados para *K. phaffii*.

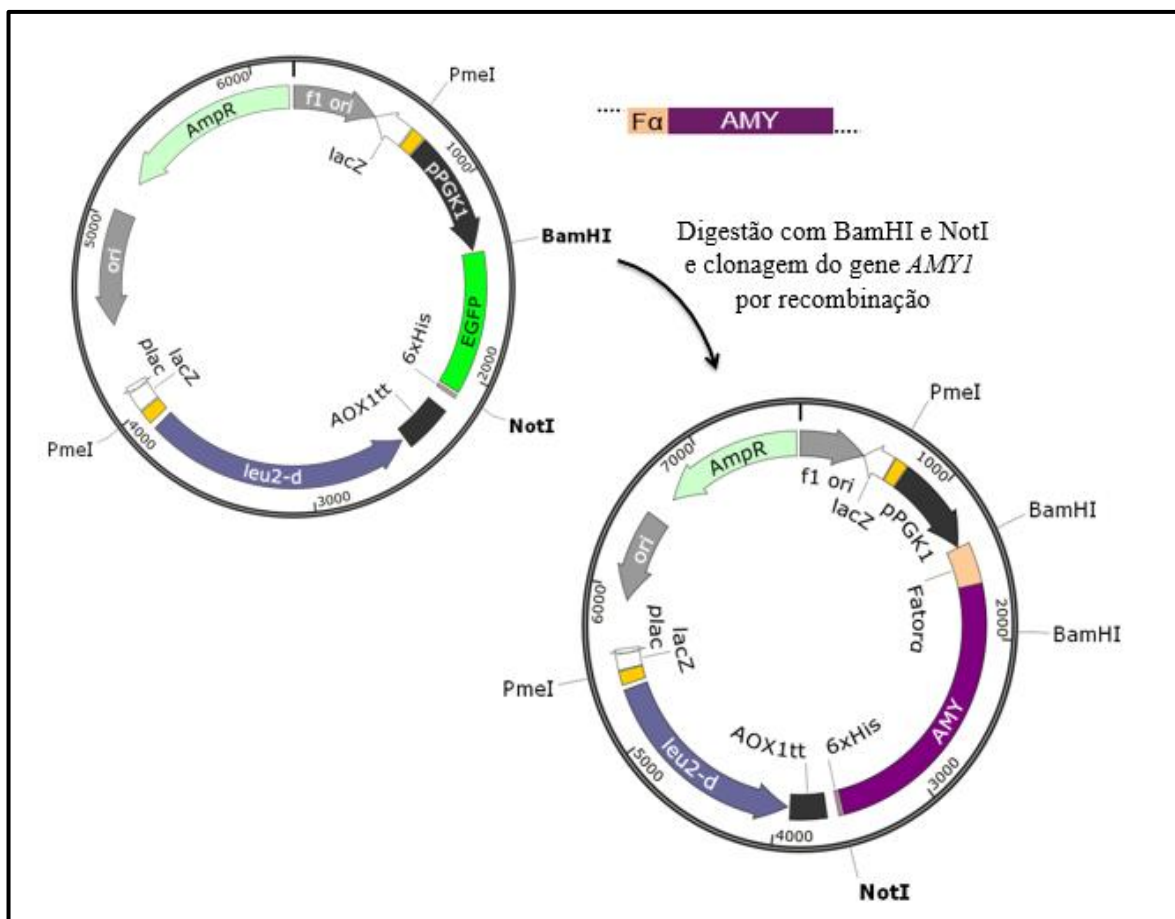
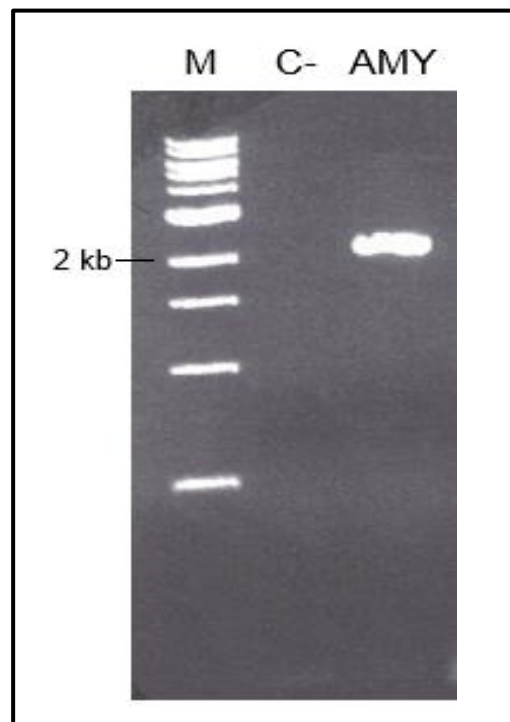


Figura 44. Estratégia para a construção dos vetores pBSChr2ld-AMY, pBSChr3ld-AMY, pBS5Sld-AMY e pPCVNTSld-AMY. Os vetores finais contêm a marca *leu2-d* e o gene *AMY1* sob controle do promotor *PGK1*.

O gene *AMY1* fusionado ao fator  $\alpha$  foi obtido por PCR a partir do vetor pNTSAMY-GCW (previamente desenvolvido no nosso laboratório) usando os *primers* Falphaif-F e AMYif-R que possuem 15 pb de homologia com o vetor (tabela 2). O fragmento amplificado de ~2,2 kb (figura 45) foi clonado por recombinação homóloga nos vetores pBSChr2ld-GFP, pBSChr3ld-GFP, pBS5Sld-GFP e pPCVNTSld-GFP usando o kit In-fusion.

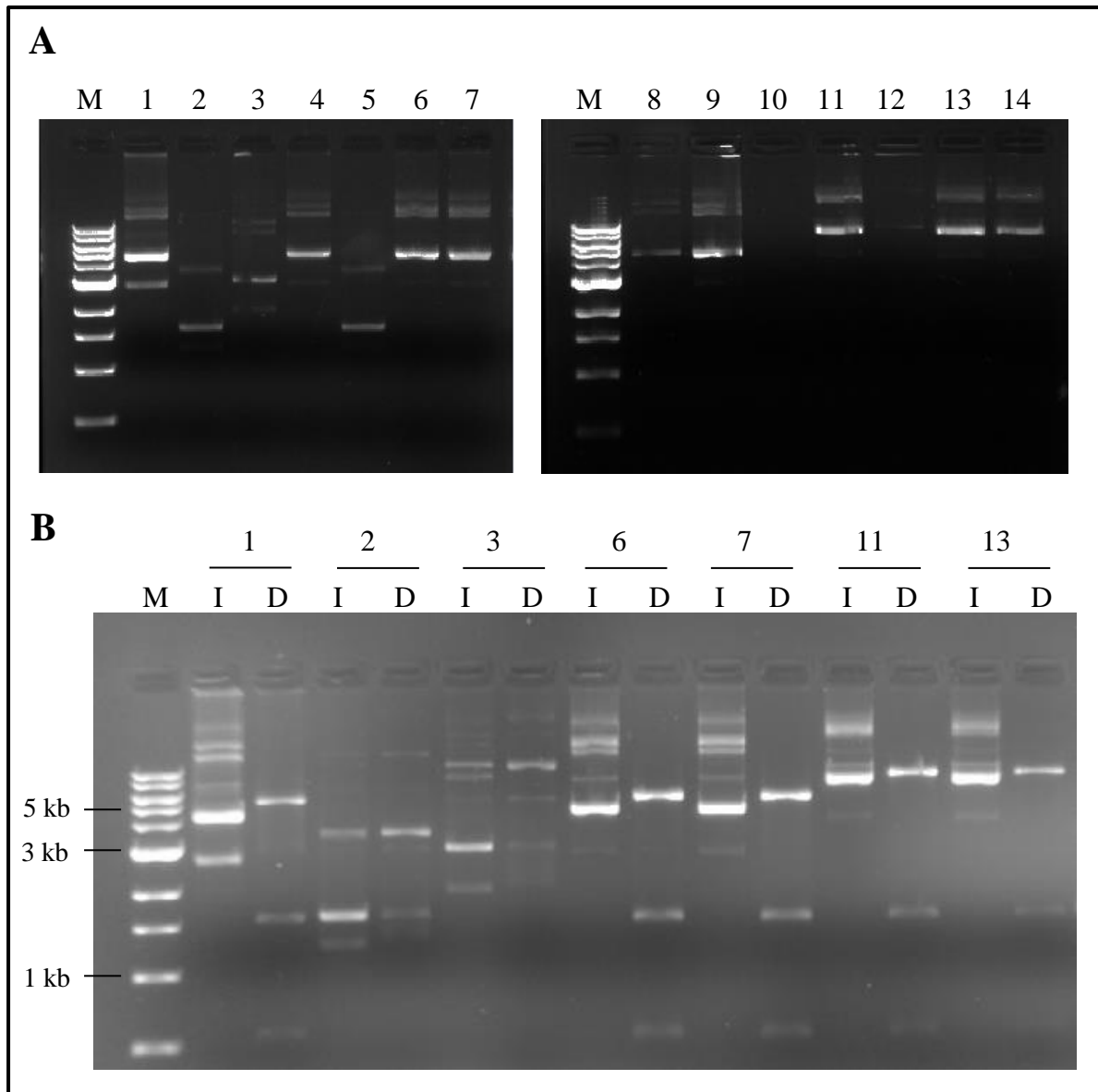


**Figura 45. Amplificação do gene *AMY1*.** Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando o resultado da PCR. M: marcador 1 kb DNA ladder (New England Biolabs), C-: controle negativo (reação sem DNA molde), AMY: reação com o vetor pNTSAMY-GCW como template.

O fragmento fator  $\alpha$  + *AMY1* também foi clonado por recombinação no vetor pKL2-GFP (igual ao vetor pKld-GFP, mas contendo a marca *LEU2* em lugar do alelo *leu2-d*) previamente digerido com BamHI e NotI. Este vetor foi construído com o fim de obter um clone com uma única cópia do cassete de expressão integrada.

Células de *E. coli* XL10-gold foram transformadas com os cinco sistemas e a seleção se deu em meio LB contendo ampicilina no caso dos vetores contendo as sequências repetidas, e em kanamicina, no caso do vetor contendo a marca *LEU2*. A figura 46A apresenta o DNA plasmidial de cinco clones transformantes dos sistemas pBSChr2ld-AMY e pBSChr3ld-AMY além de quatro clones do sistema pPCVNTSld-AMY. Os clones 1, 2 e 3 do

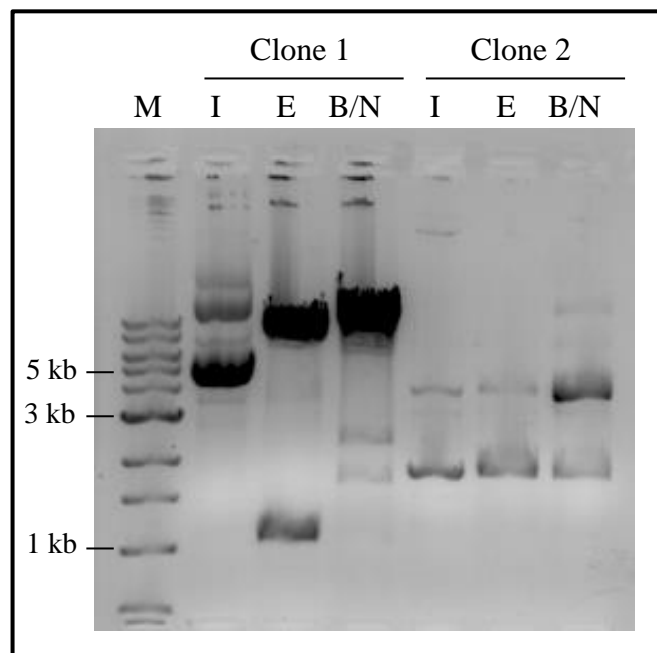
sistema pBSChr2ld-AMY (poços 1, 2 e 3), os clones 1 e 2 do sistema pBSChr3ld-AMY (poços 6 e 7) e os clones 1 e 3 do sistema pPCVNTSld-AMY (poços 11 e 13) foram selecionados para análise por digestão com as enzimas BamHI e NotI já que seu DNA plasmidial apresentou perfis diferentes. O resultado da digestão está apresentado na figura 46B.



**Figura 46. Clonagem do gene *AMY1* nos vetores com sequências repetidas.** Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando **A** DNA plasmidial para clones transformantes selecionados de cada sistema, **B** digestões com BamHI e NotI. M: marcador 1 kb DNA ladder (New England Biolabs). I: plasmídeo intacto, D: plasmídeo digerido, Poços 1-5: células transformadas com o sistema pBSChr2ld-AMY, poços 6-10: células transformadas com o sistema pBSChr3ld-AMY, poços 11-14: células transformadas com o sistema pPCVNTSld-AMY.

O clone 1 do sistema pBSChr2ld-AMY (poço 1) e os clones 1 e 2 do sistema pBSChr3ld-AMY (poços 6 e 7) apresentaram os três fragmentos esperados de aproximadamente 550 pb, 1,6 kb e 5,6 kb. Os clones 1 e 3 do sistema pPCVNTSld-AMY (poços 11 e 13) apresentaram os três fragmentos esperados de aproximadamente 550 pb, 1,6 kb e 8,6 kb.

Dois clones transformados com o vetor pBS5Sld-AMY foram analisados por digestão com a enzima EcoRV ou com BamHI e NotI. Como observado na figura 47, o clone 1 apresentou os fragmentos esperados de aproximadamente 1 kb e 6,7 kb após digestão com EcoRV. A digestão com as enzimas BamHI e NotI liberou os fragmentos esperados de aproximadamente 1,6 kb e 5,6 kb, porém, o fragmento de 550 pb não foi identificado, enquanto que um fragmento de aproximadamente 2,1 kb foi observado devido a digestão parcial do vetor. Estas digestões confirmaram a clonagem do gene *AMY1* no vetor contendo a sequência 5S.

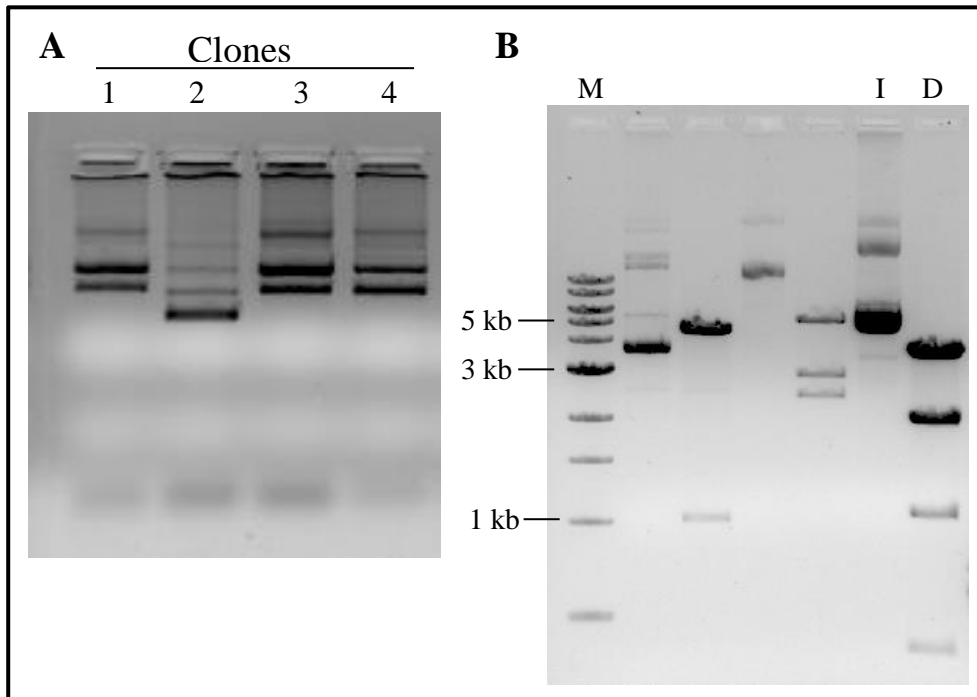


**Figura 47. Clonagem do gene *AMY1* no vetor com sequências 5S.** Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando o DNA plasmidial intacto (I), digerido com EcoRV (E) ou digerido com BamHI e NotI (B/N) para dois clones transformantes. M: marcador 1 kb DNA ladder (New England Biolabs).

A clonagem do gene *AMY1* no vetor contendo a marca *LEU2* foi confirmada por digestão com a enzima EcoRV. A figura 48A apresenta o DNA plasmidial de quatro clones



transformantes. Como observado na figura 48B, a digestão do DNA plasmidial do clone 1 liberou os fragmentos esperados de aproximadamente 400 pb, 1,1 kb, 2 kb e 3,8 kb. O vetor obtido foi chamado de pKL2-AMY.



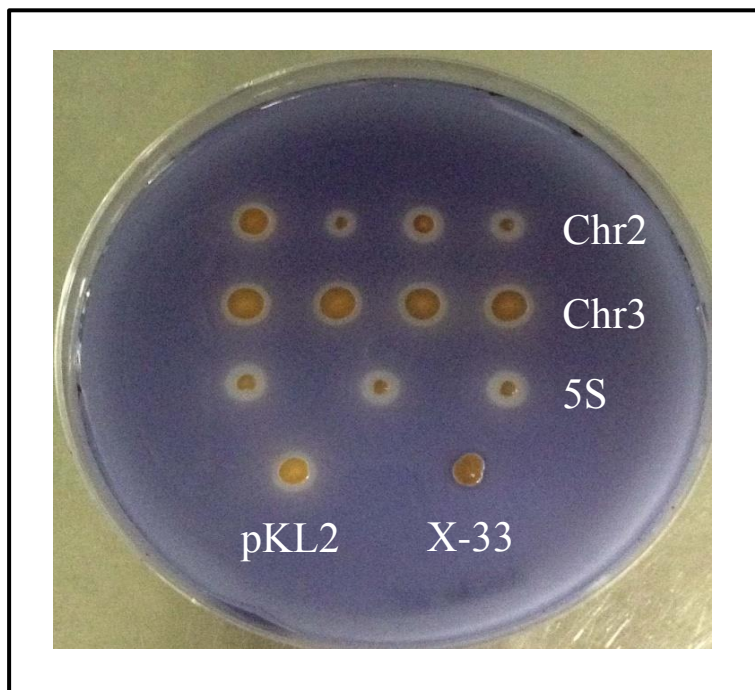
**Figura 48. Clonagem do gene *AMY1* no vetor com marca *LEU2*.** Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando **A** DNA plasmidial para quatro clones transformantes, **B** digestão com *EcoRV* do DNA plasmidial do clone 1. M: marcador 1 kb DNA ladder (New England Biolabs). I: plasmídeo intacto, D: plasmídeo digerido.

### 5.3.2 Transformação de *K. phaffii* M12

*K. phaffii* M12 foi transformada com os quatro cassetes contendo as sequências repetidas liberados após digestão dos respectivos vetores com *PmeI* e com o vetor pKL2-AMY linearizado com *SacI*. A seleção de transformantes foi feita em meio MD e nas placas foram observadas colônias de diferentes tamanhos sendo que as primeiras colônias só começaram a aparecer após quatro dias de incubação. Porém, com os cassetes contendo o gene *AMY1*, a eficiência de transformação foi muito baixa, sendo que menos de 20 colônias foram obtidas nas placas com meio MD. No caso do cassete Chr21d-AMY, apenas duas colônias cresceram na placa e houve a necessidade de se fazer uma segunda transformação para obter mais clones. No caso do cassete NTS1d-AMY, só cresceram poucas colônias, muito pequenas, após 10 dias de incubação e, portanto, este sistema não pôde ser usado para a

produção de amilase. Com o vetor pKL2-AMY a eficiência de transformação foi maior e as colônias começaram a aparecer após 2-3 dias de incubação, como normalmente acontece com células de *K. phaffii* transformadas por eletroporação com vetores contendo marcas de seleção não defectivas (Higgins e Cregg 1998).

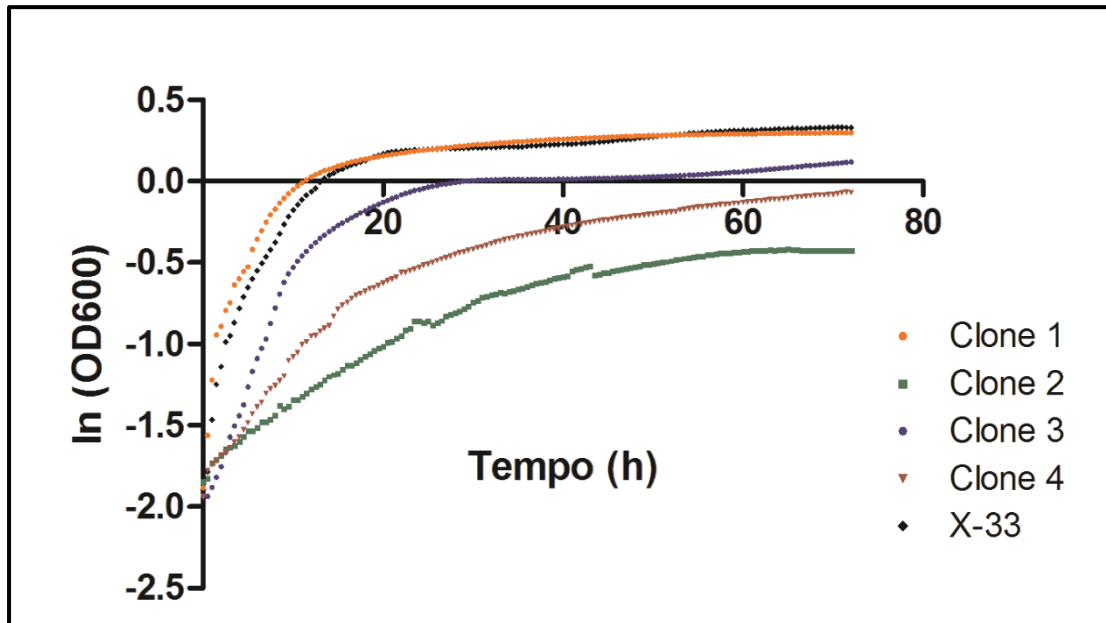
Colônias isoladas de cada sistema foram selecionadas e semeadas em meio MD contendo amido. Como controle foi semeada a levedura selvagem X-33. Após 2 dias de incubação a placa foi corada com vapor de iodo para avaliar a produção de amilase. Como observado na figura 49, a formação de halos de hidrólise do amido presente no meio constata a atividade da enzima nos clones transformantes. Apesar de o tamanho das colônias ser diferente, a razão entre o diâmetro do halo e o diâmetro da colônia foi maior para os clones transformados com os cassetes contendo as sequências repetidas em comparação ao clone transformado com o vetor pKL2-AMY. Também foi observado que os clones com menor crescimento foram os que apresentaram os maiores halos.



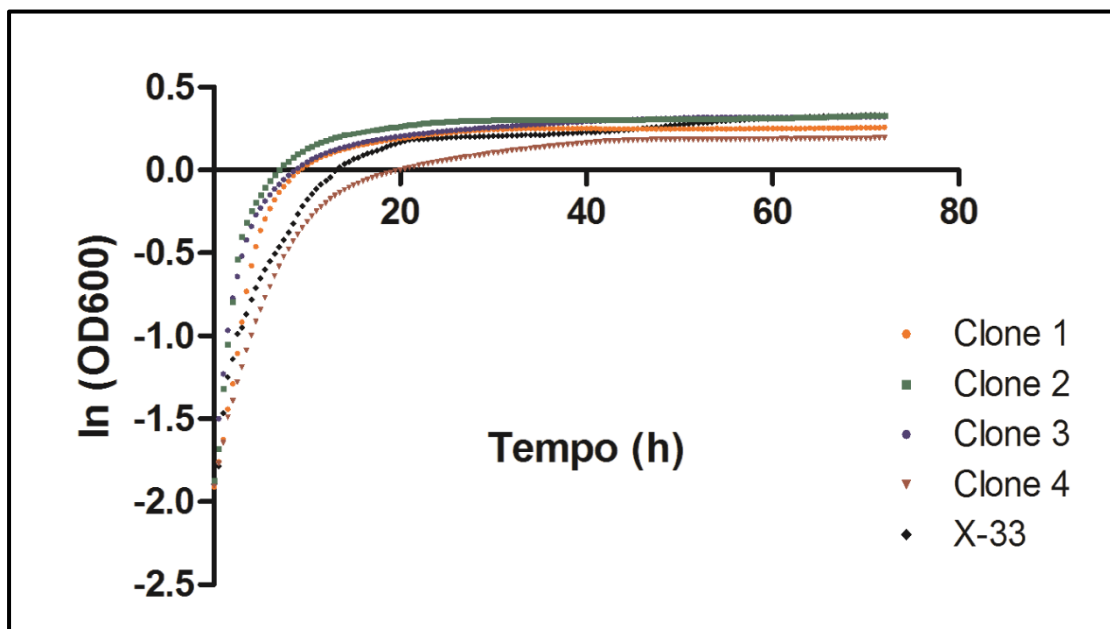
**Figura 49. Formação de halos de hidrólise de amido em clones transformados com os cassetes contendo sequências repetidas e com o vetor pKL2-AMY.** Placa com meio MD contendo amido 1% incubada por 48 h a 30°C e corada com vapor de iodo.

Quatro clones transformantes foram selecionados para os sistemas Chr2ld-AMY e Chr3ld-AMY e três clones para o sistema 5Sld-AMY para realizar cinética de crescimento em meio MD. O crescimento foi comparado com o da levedura selvagem X-33. As figuras 50, 51

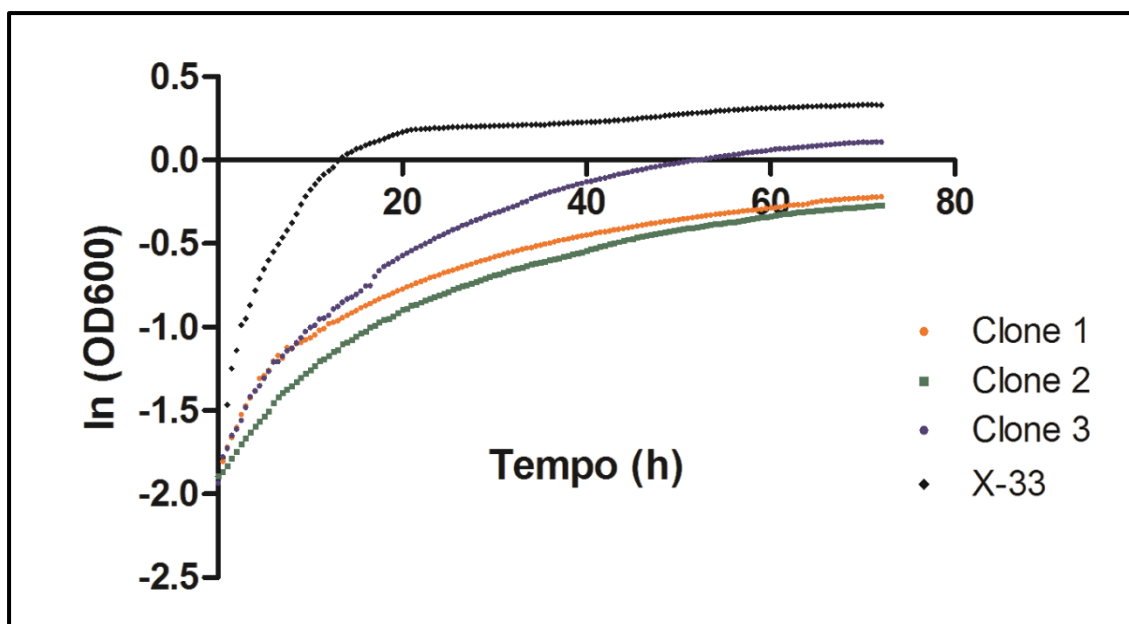
e 52 apresentam as curvas de crescimento para os clones transformados com os cassetes Chr2ld-AMY, Chr3ld-AMY e 5Sld-AMY, respectivamente. A tabela 10 apresenta a velocidade específica máxima de crescimento ( $\mu_{\text{máx}}$ ) para cada um dos clones avaliados dos quatro sistemas.



**Figura 50. Cinética de crescimento dos clones transformados com o cassete Chr2ld-AMY.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,15. X-33: linhagem selvagem de *K. phaffii*.



**Figura 51. Cinética de crescimento dos clones transformados com o cassete Chr3ld-AMY.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,15. X-33: linhagem selvagem de *K. phaffii*.



**Figura 52. Cinética de crescimento dos clones transformados com o cassete 5Sld-AMY.** Células crescidas em meio MD a 30°C durante 72 h. OD<sub>600</sub> medida a cada 30 min. OD<sub>600</sub> inicial = 0,15. X-33: linhagem selvagem de *K. phaffii*.

**Tabela 10.** Velocidade específica máxima de crescimento dos clones transformados com os cassetes Chr2ld-AMY, Chr3ld-AMY e 5Sld-AMY

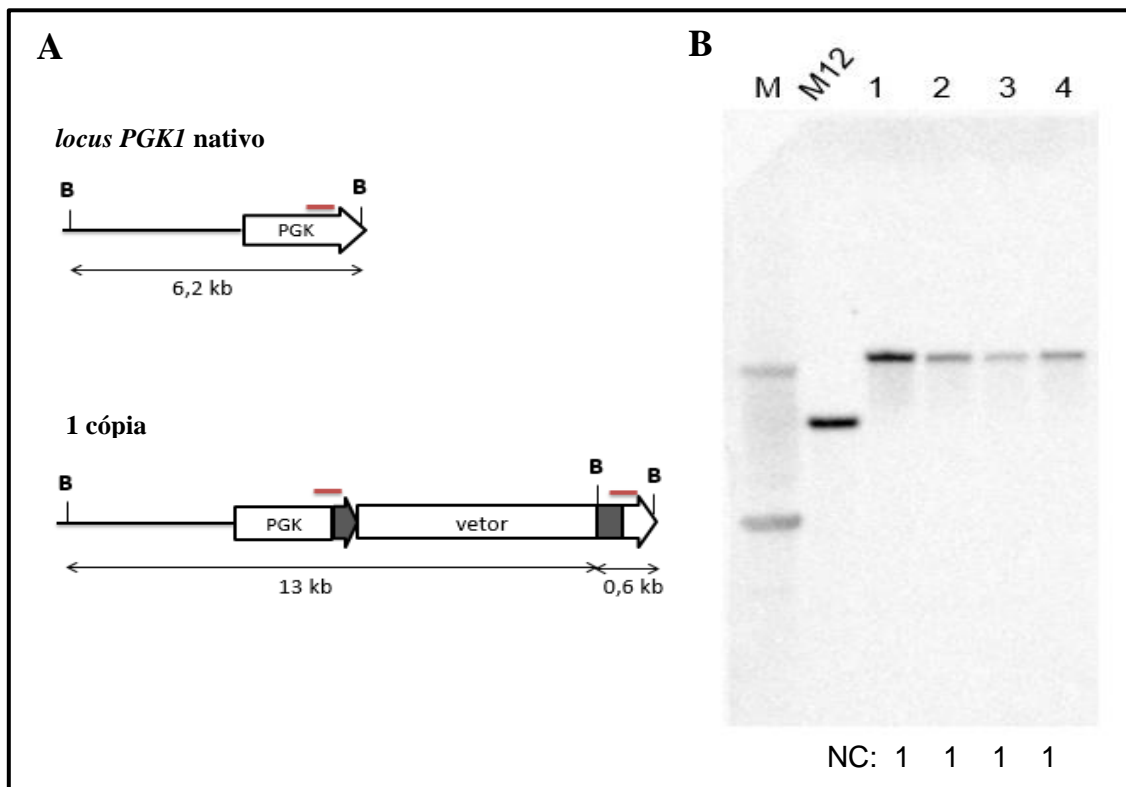
Sistema	Clone	$\mu_{\text{máx}}$ (h <sup>-1</sup> )
Chr2ld-AMY	1	0,1766 ± 0,0323
	2	0,0480 ± 0,0047*
	3	0,1520 ± 0,0096
	4	0,0766 ± 0,0087*
Chr3ld-AMY	1	0,2435 ± 0,0143
	2	0,2562 ± 0,0227
	3	0,2656 ± 0,0161
	4	0,1759 ± 0,0105
5Sld-AMY	1	0,0953 ± 0,0088*
	2	0,0701 ± 0,0020*
	3	0,0947 ± 0,0113*
X-33	-	0,2164 ± 0,0324

\*Diferença significativa em comparação à linhagem selvagem X-33 (p < 0,05). Análise de variância (ANOVA) seguida do teste de Tukey.

Os quatro clones transformados com o cassete Chr3Id-GFP, assim como os clones 1 e 3 transformados com o cassete Chr2Id-GFP, apresentaram um perfil de crescimento e uma velocidade específica máxima de crescimento semelhante ao da levedura selvagem. A velocidade específica máxima de crescimento foi menor para os outros dois clones do sistema Chr2Id-GFP e para os três clones do sistema 5SId-GFP em comparação a X-33, sendo que entre eles não houve diferença significativa.

### 5.3.3 Determinação do número de cópias

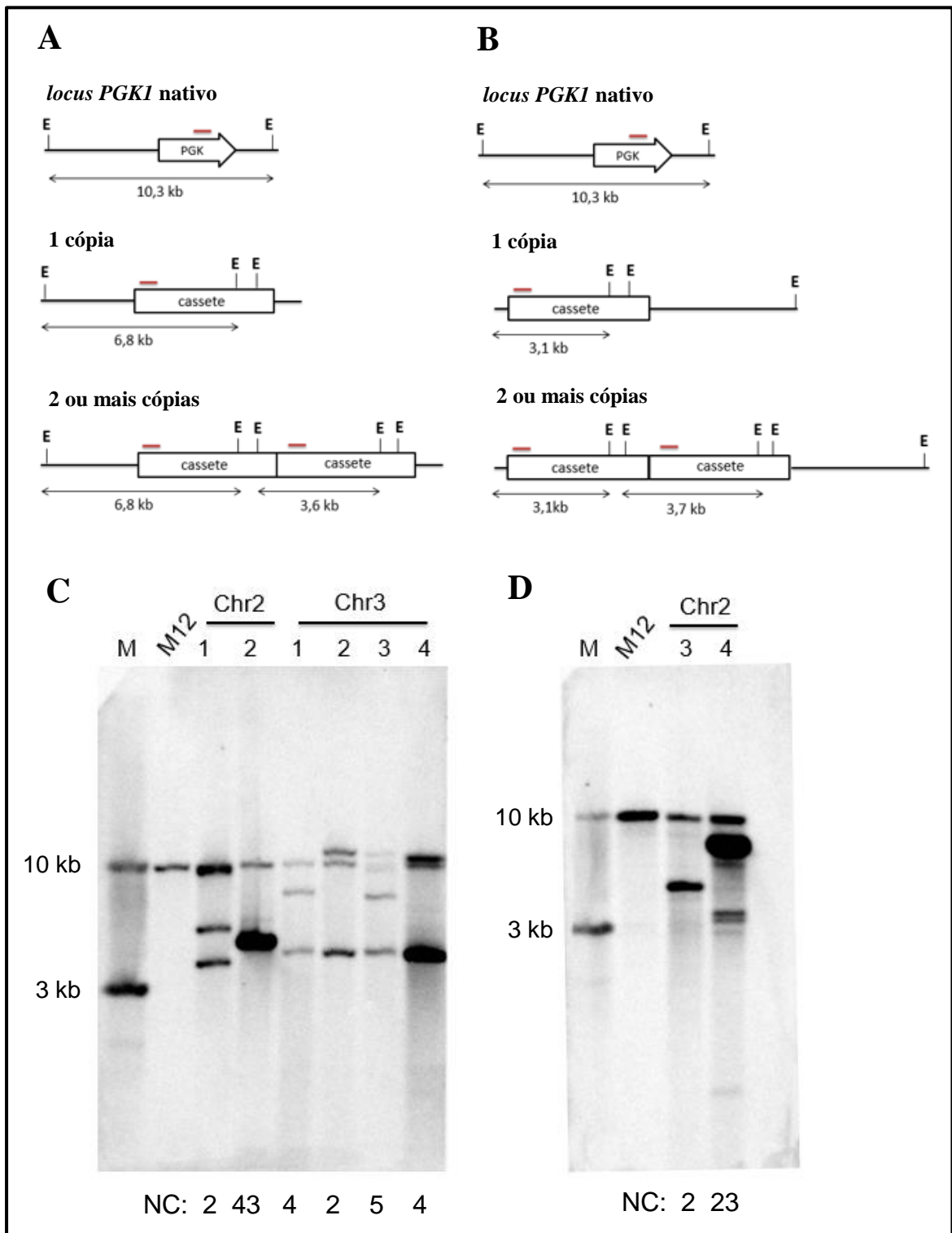
O DNA dos quatro clones transformados com o vetor pKL2-AMY foi digerido com BglII e submetido a análise por Southern blot (figura 53).



**Figura 53. Determinação do número de cópias dos clones transformados com o vetor pKL2-AMY.** A Esquema dos fragmentos esperados após digestão com BglII, B Resultado do Southern blot. B: sítio de corte para a enzima BglII, NC: número de cópias. M: marcador 1kb DNA ladder (New England Biolabs). A linha vermelha indica o lugar onde hibridiza a sonda.

A sonda utilizada, que hibridiza no promotor *PGK1*, deveria permitir a identificação de dois fragmentos de ~13 kb e ~0,6 kb se apenas uma cópia do vetor tivesse sido integrada. A figura 53A apresenta um esquema dos tamanhos dos fragmentos esperados e a figura 53B mostra o resultado do Southern blot. Como pode ser observado, todos os clones apresentaram o fragmento de ~13 kb confirmando a integração de uma cópia do vetor no *locus PGK1*. O outro fragmento de ~0,6 kb não foi observado já que pelo seu menor tamanho o mesmo percorreu o gel e saiu durante a migração eletroforética. O fragmento de ~6,2 kb observado na linhagem M12 correspondente ao *locus PGK1* intacto.

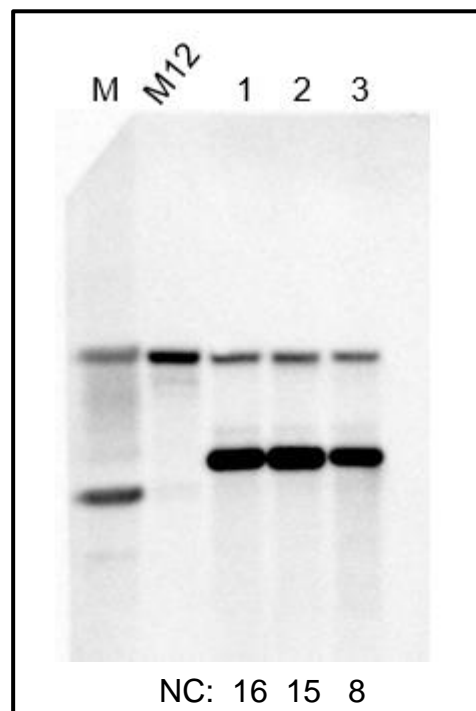
O DNA dos quatro clones transformados com o cassete Chr2ld-AMY e dos quatro clones transformados com o cassete Chr3ld-AMY foi digerido com EcoRV e foi usada a sonda que hibridiza no promotor *PGK1* para análise de Southern blot. No caso do cassete Chr2ld-AMY deveria ser identificado um fragmento de ~6,8 kb se apenas uma cópia tivesse sido integrada. Se tivessem sido integradas duas cópias do cassete, seria identificado um outro fragmento de ~3,6 kb cuja intensidade iria aumentando para cada cópia a mais integrada. No caso do cassete Chr3ld-AMY, deveria ser identificado um fragmento de ~3,1 kb se apenas uma cópia tivesse sido integrada. Se tivessem sido integradas duas cópias do cassete seria identificado um outro fragmento de ~3,7 kb cuja intensidade iria aumentando por cada cópia a mais integrada. Para ambos os casos seria observado um fragmento de ~10,3 kb correspondente ao *locus* do *PGK1* se a integração ocorresse em outro *locus* diferente. Os painéis A e B da figura 54 apresentam um esquema dos tamanhos dos fragmentos esperados para os cassetes Chr2ld-AMY e Chr3ld-AMY, respectivamente. O painel C da figura 54 mostra o resultado do Southern blot para os clones 1 e 2 transformados com o cassete Chr2ld-AMY e para os quatro clones transformados com o cassete Chr3ld-AMY. O painel D da figura 54 mostra o resultado do Southern blot para os clones 3 e 4 transformados com o cassete Chr2ld-AMY. Todos os clones de ambos os sistemas apresentaram o fragmento de ~10,3 kb indicando que a integração não ocorreu no *locus PGK1*, porém, esta pode ter ocorrido em outros *loci* diferentes aos cromossomos 2 e 3 já que foram observados outros fragmentos além dos esperados. A presença do fragmento de ~3,6 kb nos clones 1 e 4 do sistema Chr2ld-AMY e do fragmento de ~3,7 kb nos quatro clones do sistema Chr3ld-AMY confirmou a integração de múltiplas cópias *in tandem*.



**Figura 54. Determinação do número de cópias dos clones transformados com os cassetes Chr2ld-AMY e Chr3ld-AMY.** **A** Esquema dos fragmentos esperados após digestão com EcoRV para os clones do sistema Chr2ld-AMY, **B** Esquema dos fragmentos esperados após digestão com EcoRV para os clones do sistema Chr3ld-AMY, **C** Resultado do Southern blot para os clones 1 e 2 do sistema Chr2ld-AMY e para os quatro clones do sistema Chr3ld-AMY, **D** Resultado do Southern blot para os clones 3 e 4 do sistema Chr2ld-AMY. E: sítio de corte para a enzima EcoRV. NC: número de cópias. M: marcador 1kb DNA ladder (New England Biolabs). A linha vermelha indica o lugar onde hibridiza a sonda.

Usando como referência o fragmento de ~10,3 kb foi determinado o número de cópias: 2, 43, 2 e 23 cópias para os clones 1, 2, 3 e 4 do sistema Chr2ld-AMY, respectivamente e 4, 2, 5 e 4 cópias para os clones 1, 2, 3 e 4 do sistema Chr3ld-AMY, respectivamente. É importante notar que os clones com maior número de cópias (43 e 23) foram os que apresentaram menor taxa de crescimento, como tinha sido observado com o cassete Chr2ld-GFP.

O DNA dos três clones transformados com o cassete 5Sld-AMY foi digerido com EcoRV (figura 55). Com a sonda que hibridiza no promotor *PGK1* foi identificado um fragmento de ~3,7 kb que confirmou a integração de múltiplas cópias *in tandem* em um único *locus*, como tinha sido observado com o cassete 5Sld-GFP. Todos os clones apresentaram o fragmento de ~10,3 kb indicando que a integração não ocorreu no *locus PGK1* e este fragmento foi usado como referência para determinar o número de cópias: 16, 15 e 8 cópias para os clones 1, 2 e 3, respectivamente. Resultados similares tinham sido observados em *S. cerevisiae* quando foram usadas como alvo de integração as sequências  $\delta$  que se encontram espalhadas no genoma da levedura (Wang *et al.*, 1996). Apesar de existirem aproximadamente 100 cópias dessas sequências, as múltiplas cópias dos vetores de expressão foram integradas *in tandem* no máximo em três sítios diferentes.



**Figura 55. Determinação do número de cópias dos clones transformados com o cassete 5Sld-AMY.** Resultado do Southern blot após digestão do gDNA com EcoRV. NC: número de cópias. M: marcador 1kb DNA ladder (New England Biolabs).



Devido ao fato de que a taxa de transcrição das marcas auxotróficas com promotores truncados ser baixa, era esperado que, para recuperar a prototrofia, várias cópias de vetor iriam ser integradas no genoma de levedura. Porém, os resultados deste trabalho mostraram que apenas duas cópias do vetor contendo a marca *leu2-d* foram necessárias para que *K. phaffii* M12 pudesse crescer em meio mínimo. Em *S. cerevisiae* havia sido observado que de duas a cinco cópias da marca *leu2-d* são necessárias para produzir suficiente leucina para a célula (Lopes *et al.*, 1991).

Ao relacionar o número de cópias com a velocidade específica máxima de crescimento, foi observada uma correlação negativa como mostrado na figura 56.

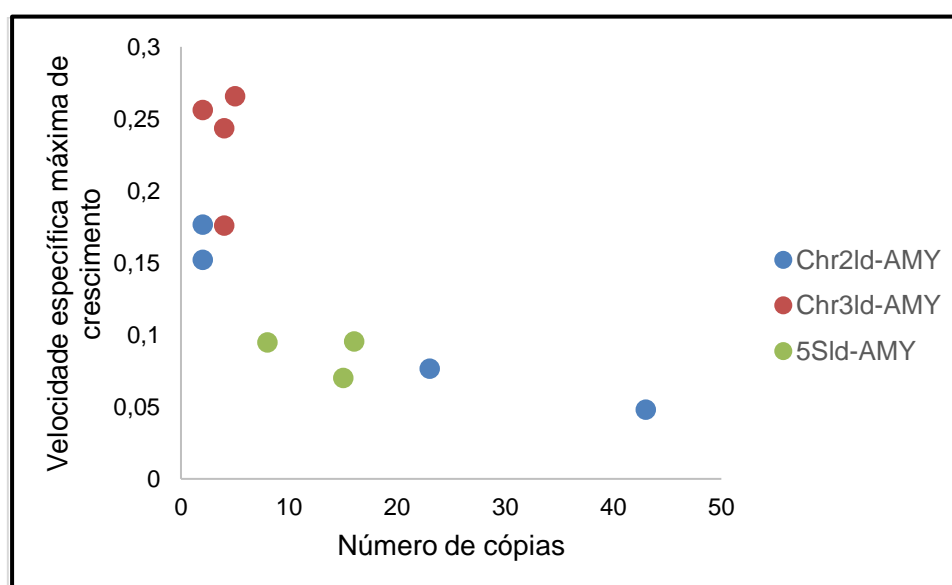


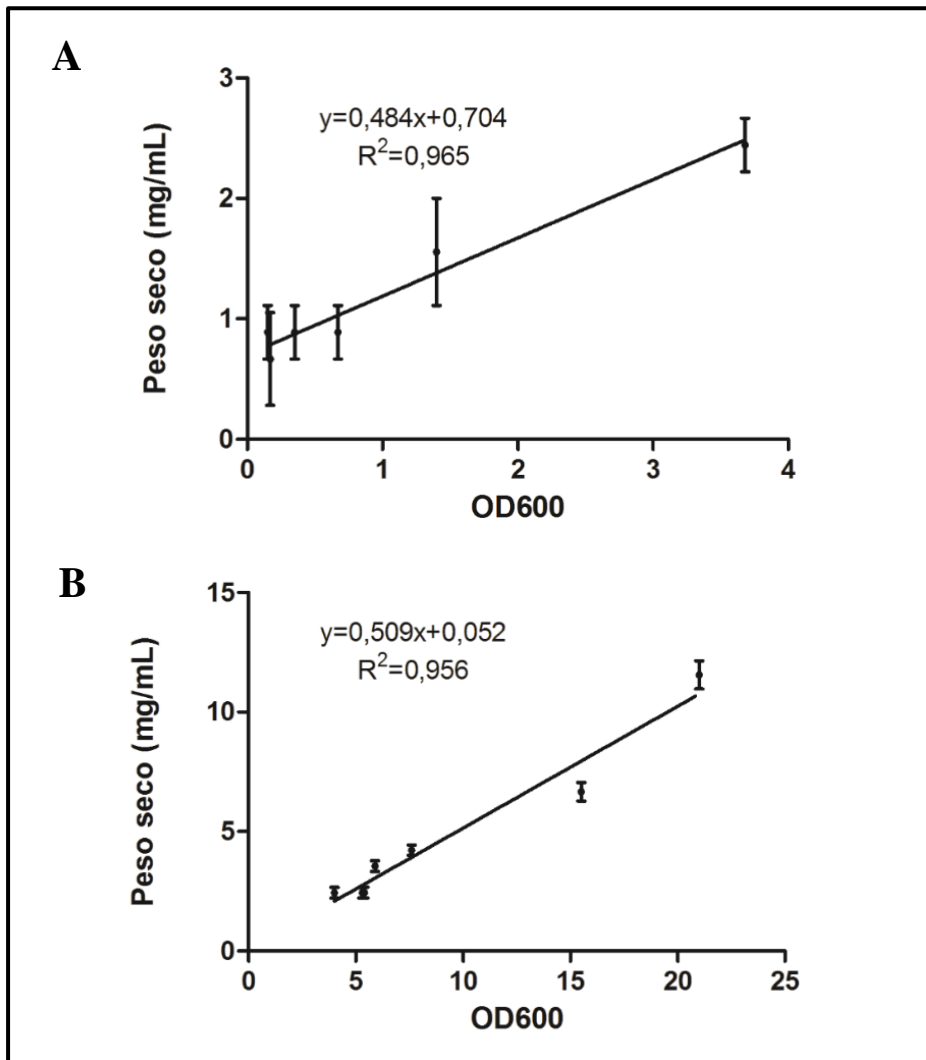
Figura 56. Correlação entre a velocidade específica máxima de crescimento e o número de cópias dos clones transformados com os cassetes contendo sequências repetidas e o gene *AMY1*.

#### 5.3.4 Produção de amilase

Existem vários métodos para determinar a atividade da  $\alpha$ -amilase, baseados principalmente no aumento de açúcares redutores ou na diminuição na intensidade da cor formada pelo complexo iodo-amido. Quando uma solução de amido é misturada com iodo/iodeto forma-se o complexo iodo-amido que produz uma cor azul intensa. Esta cor é devida principalmente à interação de moléculas de iodo carregadas com a hélice formada pela amilose. Esta interação forma complexos poliiodetos que apresentam um pico máximo de absorvância a um comprimento de onda de aproximadamente 650 nm (Thoma e French

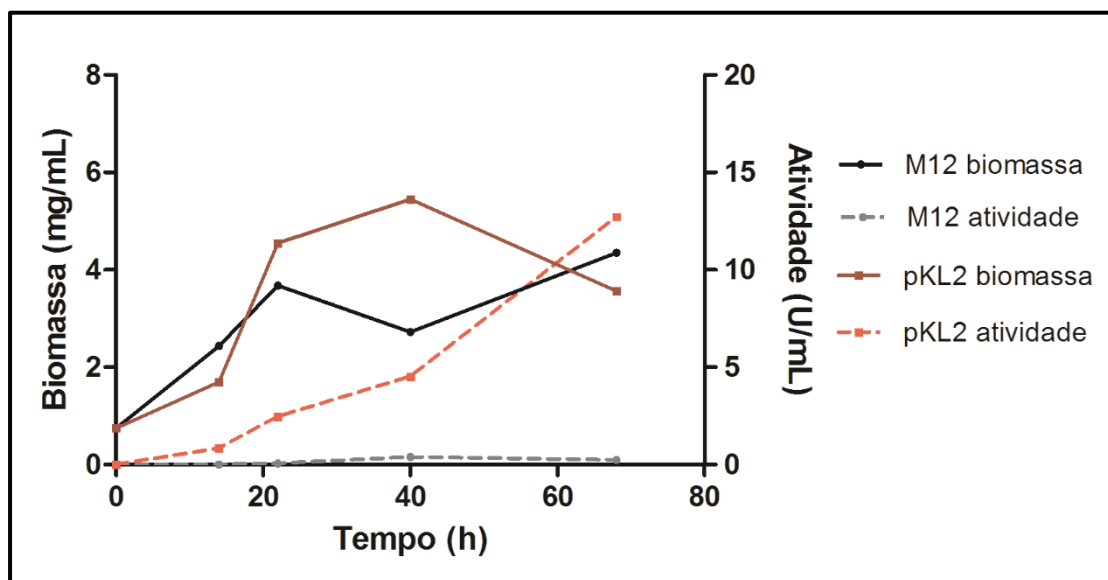
1960). Desta forma, métodos colorimétricos podem ser usados para determinar a atividade de amilases, baseados na mudança da intensidade da cor do complexo iodo-amido conforme o amido é hidrolisado.

Três clones de cada sistema foram avaliados para a produção de amilase usando o método colorimétrico descrito por Fuwa (1954). Para o sistema Chr2ld-AMY, foram selecionados os clones 2, 3 e 4, para o sistema Chr3ld-AMY, os clones 1, 3 e 4. Durante o ensaio, foi acompanhado o crescimento celular com a determinação da OD<sub>600</sub> e foram usadas as correlações apresentadas na figura 57 para determinar a biomassa (mg/mL). A correlação do painel A corresponde a valores de OD<sub>600</sub> entre 0,15 e 3,7 e a correlação do painel B corresponde a valores de OD<sub>600</sub> entre 4 e 21.



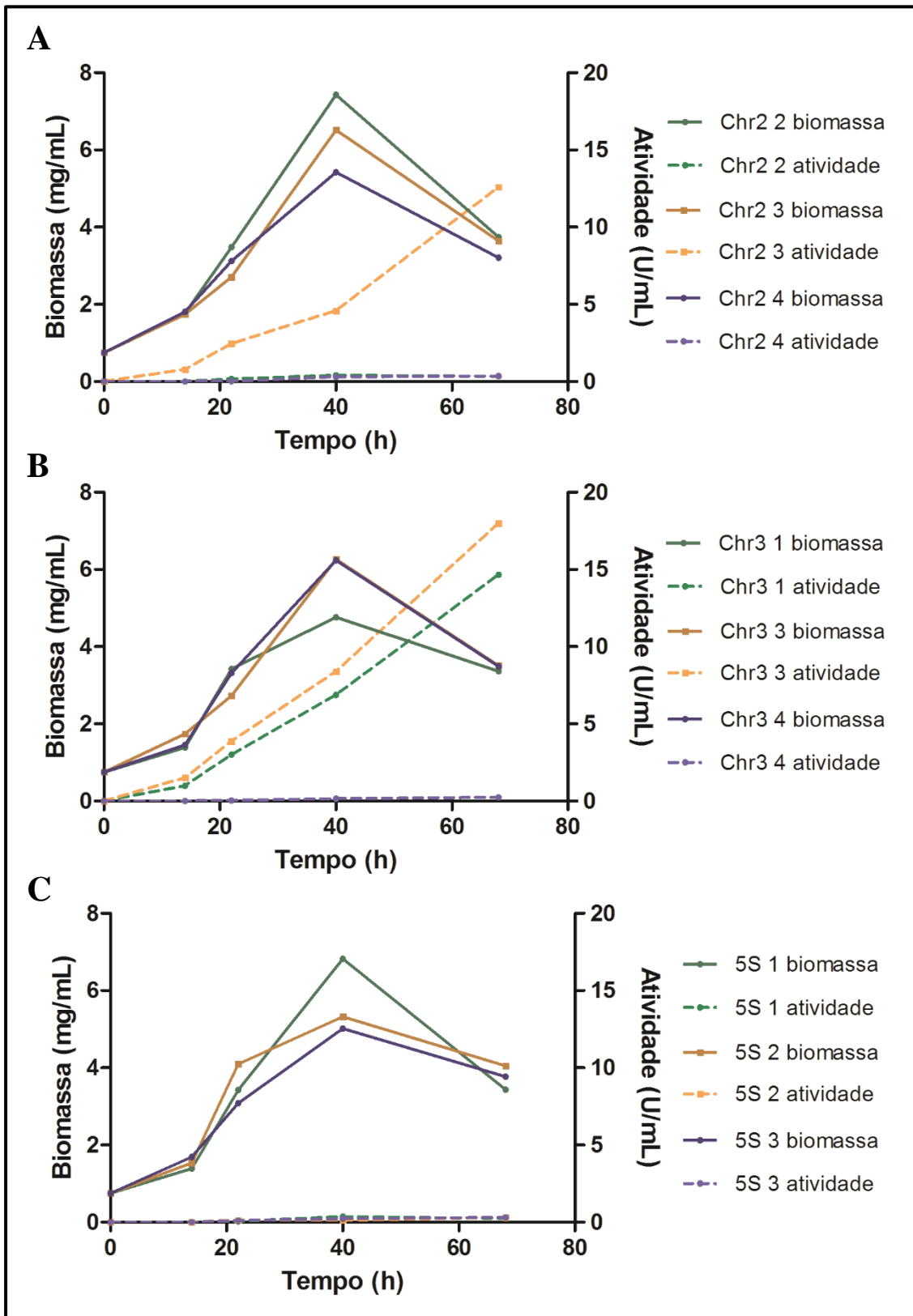
**Figura 57. Correlação entre o peso seco e a OD<sub>600</sub> da levedura *K. phaffii* M12. A Correlação para valores de OD<sub>600</sub> entre 0,15 e 3,7. B Correlação para valores de OD<sub>600</sub> entre 4 e 21. As barras de erro representam o erro padrão da média para n=3.**

Inicialmente, o ensaio foi realizado em meio complexo e, para isso, os clones foram crescidos em meio YPD durante 68 h e a atividade amilásica no sobrenadante foi determinada nos tempos de 14 h, 22 h, 40 h e 68 h. A figura 58 apresenta o resultado do ensaio de atividade assim como a cinética de crescimento para a linhagem *K. phaffii* M12 e para o clone contendo uma cópia do gene *AMY1*. Pode ser observado que a levedura selvagem não apresentou atividade amilolítica durante todo o experimento enquanto que a atividade no clone com o vetor pKL2-AMY integrado no genoma aumentou com o crescimento da levedura.



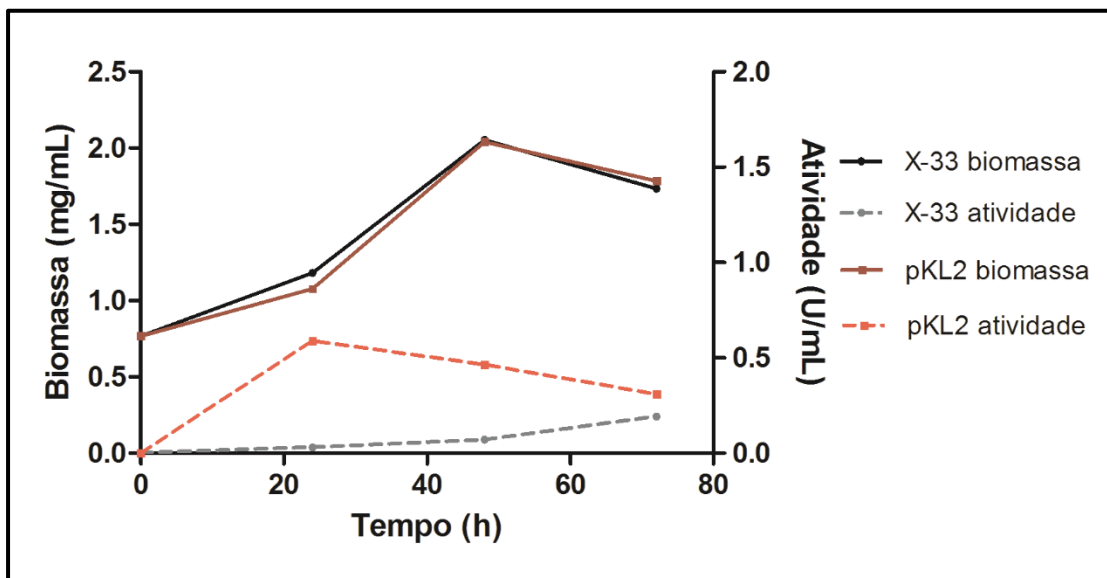
**Figura 58.** Cinética de produção de amilase em meio complexo na linhagem *K. phaffii* M12 e no clone contendo uma cópia do gene *AMY1*. Células crescidas em meio YPD a 28°C e 250 rpm durante 68 h. Atividade medida no sobrenadante.

A figura 59 apresenta as cinéticas de crescimento e de produção de amilase em meio YPD para os clones transformados com os cassetes contendo as sequências repetidas. O painel A mostra os resultados dos clones do sistema Chr2ld-AMY, o painel B os resultados dos clones do sistema Chr3ld-AMY e o painel C, os resultados dos clones do sistema 5Sld-AMY. Como observado nos gráficos só foi detectada atividade amilolítica no clone 3 do sistema Chr2ld-AMY e nos clones 1 e 3 do sistema Chr3ld-AMY. Essa atividade aumentou com o aumento da biomassa.



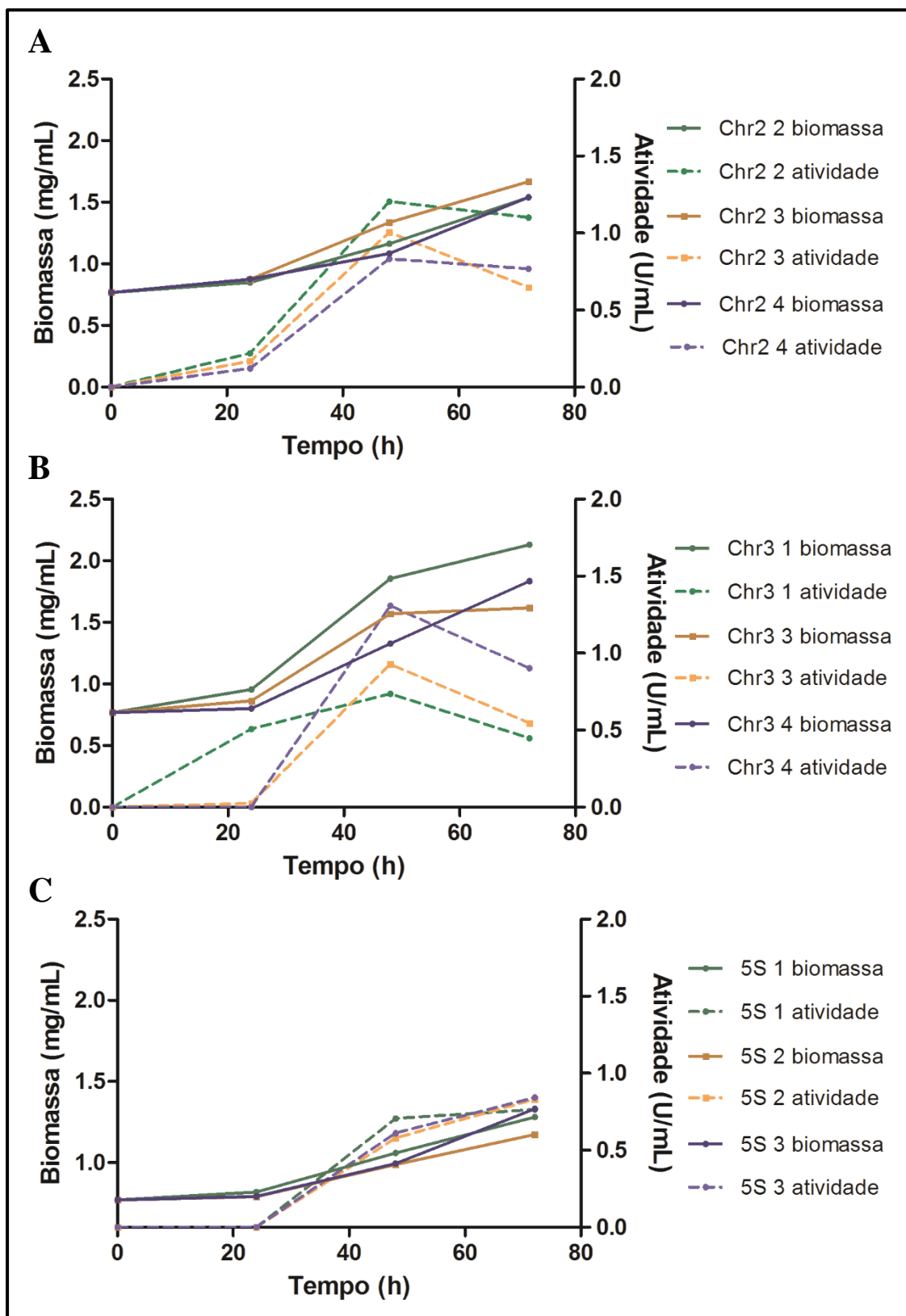
**Figura 59. Cinética de produção de amilase em meio complexo nos clones transformados com os cassetes contendo seqüências repetidas.** Células crescidas em meio YPD a 28°C e 250 rpm durante 68 h. **A** Sistema Chr2ld-AMY, **B** Sistema Chr3ld-AMY, **C** Sistema 5Sld-AMY. Atividade medida no sobrenadante.

A ausência de atividade nos demais clones transformados com os cassetes contendo as sequências repetidas pode ser devida à perda das cópias do cassete por falta de pressão seletiva no meio complexo e, portanto, foi feito o ensaio em meio mínimo. Os clones foram crescidos em meio MD durante 72 h e a atividade da amilase no sobrenadante foi determinada a cada 24 h. A figura 60 apresenta a cinética de crescimento e de produção de amilase para a linhagem *K. phaffii* X-33 e para o clone contendo uma cópia do gene *AMY1*. Como observado anteriormente no meio complexo, a levedura selvagem não apresentou atividade amilolítica durante todo o experimento em meio mínimo, enquanto que no clone com o vetor pKL2-AMY integrado no genoma foi observada atividade com um pico máximo após 24 h de cultivo.



**Figura 60.** Cinética de produção de amilase em meio mínimo na linhagem *K. phaffii* X-33 e no clone contendo uma cópia do gene *AMY1*. Células crescidas em meio MD a 28°C e 250 rpm durante 72 h. Atividade medida no sobrenadante.

A figura 61 apresenta as cinéticas de crescimento e de produção de amilase em meio MD para os clones transformados com os cassetes contendo as sequências repetidas. O painel A mostra os resultados dos clones do sistema Chr2ld-AMY, o painel B, os resultados dos clones do sistema Chr3ld-AMY e, o painel C, os resultados dos clones do sistema 5Sld-AMY. Usando meio mínimo para promover o crescimento da levedura foi possível detectar atividade amilolítica em todos os clones. Essa atividade aumentou com o aumento da biomassa, sendo que o valor máximo foi detectado após 48 h de cultivo.



**Figura 61. Cinética de produção de amilase em meio mínimo nos clones transformados com os cassetes contendo sequências repetidas.** Células crescidas em meio MD a 28°C e 250 rpm durante 72 h. **A** Sistema Chr2ld-AMY, **B** Sistema Chr3ld-AMY, **C** Sistema 5Sld-AMY. Atividade medida no sobrenadante.

Quando crescida em meio mínimo, *K. phaffii* atingiu valores de biomassa e de atividade muito menores aos obtidos no meio complexo, devido provavelmente a que parte da fonte de carbono no meio mínimo (neste caso glicose) é utilizada para sintetizar os aminoácidos que não estão presentes no meio de cultura. Desta forma, após 48 h de cultivo, a pouca glicose presente no meio deve ser usada pela levedura para a manutenção celular e não para a produção da proteína heteróloga, o que explicaria a diminuição da atividade enzimática nesse ponto da curva de produção.

A tabela 11 apresenta os valores de atividade (U/mL) e de atividade relacionada com a concentração celular (U/mg) para os clones analisados em meio mínimo. Todos os clones contendo múltiplas cópias do gene *AMY1* apresentaram valores de atividade maiores aos apresentados pelo clone contendo uma única cópia do gene, confirmando que a integração múltipla aumenta a produção da proteína heteróloga.

**Tabela 11.** Atividade amilolítica no sobrenadante dos clones transformados com os cassetes contendo sequências repetidas crescidos em meio mínimo

Clone	Atividade (U/mL)				Atividade (U/mg)			
	0 h	24 h	48 h	72 h	0 h	24 h	48 h	72 h
X-33	0,000	0,032	0,070	0,193	0,000	0,027	0,034	0,111
pKL2	0,000	0,589	0,464	0,308	0,000	0,546	0,227	0,173
Chr2 2	0,000	0,218	1,204	1,100	0,000	0,258	1,035	0,714
Chr2 3	0,000	0,168	1,004	0,648	0,000	0,191	0,751	0,389
Chr2 4	0,000	0,120	0,832	0,767	0,000	0,137	0,768	0,498
Chr3 1	0,000	0,508	0,736	0,447	0,000	0,532	0,397	0,210
Chr3 3	0,000	0,024	0,928	0,546	0,000	0,027	0,591	0,338
Chr3 4	0,000	0,000	1,308	0,902	0,000	0,000	0,984	0,492
5S 1	0,000	0,000	0,706	0,764	0,000	0,000	0,667	0,596
5S 2	0,000	0,000	0,579	0,831	0,000	0,000	0,587	0,710
5S 3	0,000	0,000	0,610	0,843	0,000	0,000	0,615	0,633

A tabela 12 resume os valores de produção relativa de amilase assim como o número de cópias integradas do cassete e a velocidade específica máxima de crescimento para cada clone analisado.

**Tabela 12.** Relação de número de cópias, produção relativa de amilase e velocidade de crescimento

Sistema	Clone	Número de cópias	Produção relativa de amilase*	$\mu_{\text{máx}}$
	controle	1	1	
Chr2ld-AMY	2	43	2,6	0,0480
	3	2	3,3	0,1520
	4	23	3,4	0,0766
Chr3ld-AMY	1	4	1,7	0,2435
	3	5	2,6	0,2656
	4	4	4,3	0,1759
5Sld-AMY	1	16	2,9	0,0953
	2	15	2,6	0,0701
	3	8	2,7	0,0947

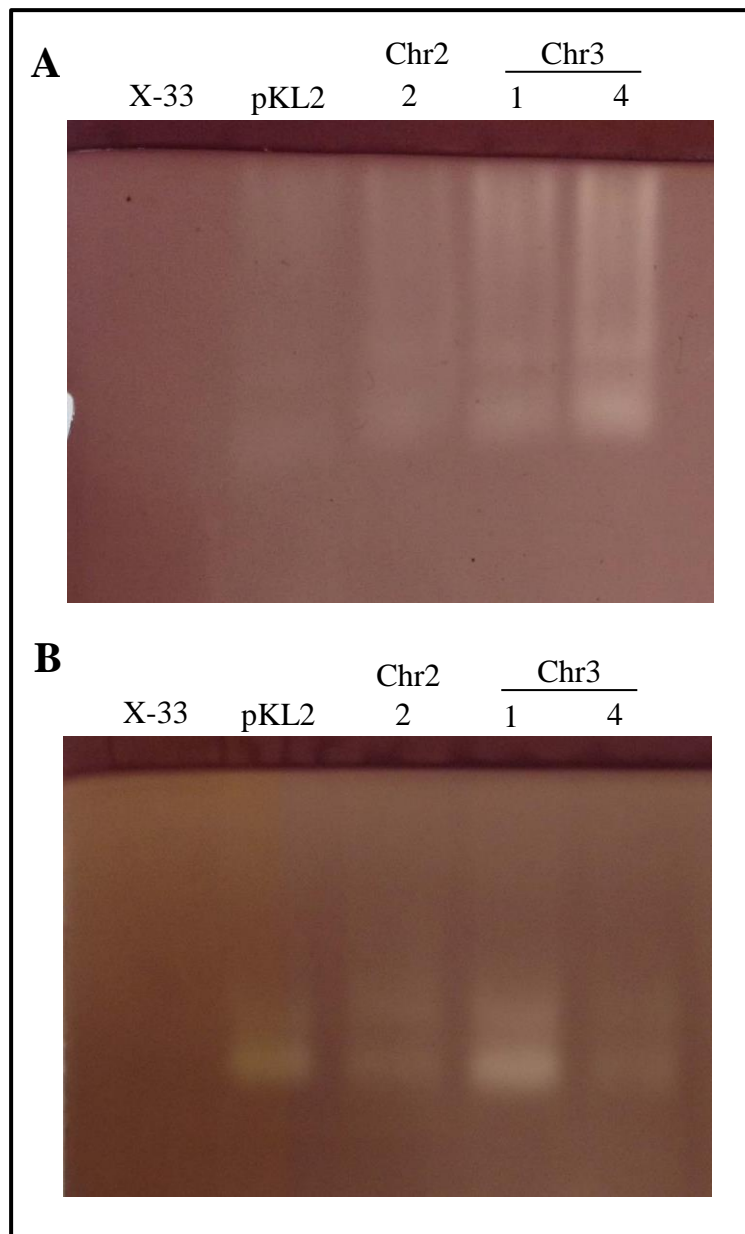
\* Produção em relação ao clone com uma cópia do gene.

É importante notar que clones contendo o mesmo número de cópias, como os clones 1 e 4 do sistema Chr3ld-AMY (4 cópias), não apresentaram a mesma atividade da amilase. A atividade por mg de célula foi muito maior no clone 4 em comparação ao clone 1. Isto pode ser devido à integração dos cassetes de expressão em locais diferentes e, no caso do clone 1, esta pode ter ocorrido em um local de baixa expressão gênica. Além disso, o clone 1 apresentou uma taxa de crescimento maior em comparação ao clone 4 o que poderia sugerir que a maior produção de amilase no clone 4 tem um efeito negativo no crescimento da célula.

O clone 2 do sistema Chr2ld-AMY e os clones 1 e 4 do sistema Chr3ld-AMY foram selecionados e crescidos em meio mínimo com glicerol como fonte de carbono já que consideramos que a glicose poderia estar inibindo a atividade enzimática (inibição por substrato). Como controle, foram crescidas células de *K. phaffii* X-33 e o clone com uma cópia do gene *AMY1*. Após 48 h de incubação, foi coletado o sobrenadante e as células foram rompidas para obter o extrato proteico intracelular. Ambas as amostras foram aplicadas em gel não-desnaturante e foi feito um zimograma. Como era esperado, X-33 não apresentou atividade, nem no sobrenadante nem dentro da célula (figura 62). Como pode ser observado no painel A da figura 62, para os três clones avaliados, parte da amilase ficou retida dentro da célula, o que explicaria os baixos níveis de atividade amilolítica e a baixa concentração da



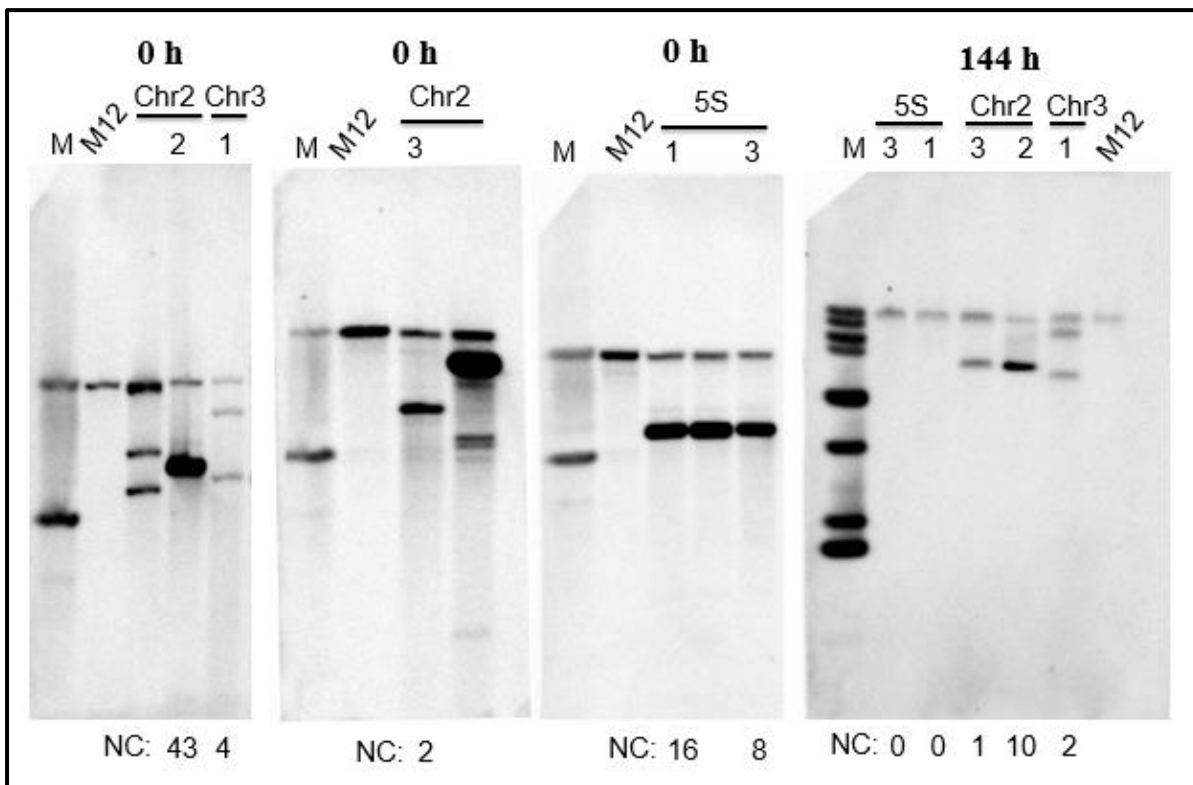
proteína no sobrenadante (figura 62B). Contudo, este resultado mostra que o gene *AMY1* está sendo expresso em maior quantidade nos clones com múltiplas cópias e a produção da proteína poderia ser otimizada melhorando o sistema de secreção da célula.



**Figura 62. Zimograma de clones selecionados produzindo amilase.** Gel SDS-PAGE 10%. Clones crescidos em meio MD com glicerol durante 48 h a 28°C e 250 rpm. **A** extrato proteico intracelular, **B** sobrenadante. X-33: linhagem selvagem de *K. phaffii*. pKL2: clone contendo 1 cópia do gene *AMY1*.

### 5.3.5 Estabilidade genética

Os clones 2 e 3 do sistema Chr2ld-AMY, o clone 1 do sistema Chr3ld-AMY e os clones 1 e 3 do sistema 5Sld-AMY foram selecionados e crescidos em meio não-seletivo (YPD) durante 72 gerações para avaliar a estabilidade genética. A seleção desses clones foi feita de modo a ter um clone com alto número de cópias e outro com baixo número de cópias para cada sistema. Como todos os clones de sistema Chr3ld-AMY tinham número similar de cópias, foi analisado um único clone. A cultura foi transferida para meio fresco a cada 24 h para garantir que as células estivessem na fase exponencial durante todo o experimento. Após 144 h de crescimento, o número de cópias foi analisado por Southern blot após digestão do DNA genômico com a enzima EcoRV. Como observado na figura 63, todos os clones avaliados apresentaram baixa estabilidade genética.



**Figura 63.** Estabilidade genética dos clones transformados com os cassetes Chr2ld-AMY, Chr3ld-AMY e 5Sld-AMY. Resultado do Southern blot antes (0 h) e após 144 h de crescimento em meio não seletivo (YPD). NC: número de cópias. M: marcador 1kb DNA ladder (New England Biolabs) nos géis de 0 h ou 2-log DNA ladder (New England Biolabs) no gel de 144 h.

O clone 3 do sistema Chr2ld-AMY e o clone 1 do sistema Chr3ld-AMY que inicialmente tinham baixo número de cópias perderam 50% das cópias integradas ficando com 1 e 2 cópias, respectivamente. O clone 2 do sistema Chr2ld-AMY, que tinha inicialmente 43 cópias, perdeu 77% das cópias integradas ficando com apenas 10 cópias do cassete. Os dois clones do sistema 5Sld-AMY perderam 100% das cópias do cassete. A diminuição do número de cópias ou a perda total do cassete está em concordância com a falta de atividade amilolítica nesses clones quando crescidos em meio YPD. Estes resultados confirmam que alto números de cópias integradas *in tandem* no *locus* do cromossomo 2 são mais propensos a sofrer recombinação homóloga intramolecular o que leva a perder uma alta porcentagem do número de cópias inicial como foi observado também com o cassete Chr2ld-GFP. Igualmente, este teste de estabilidade confirmou que o *locus* 5S é uma região altamente instável quando não é mantida a pressão seletiva.

#### 5.4 Correlação entre o número de cópias e a produção de proteínas

Depois de ter confirmado que os diversos sistemas testados neste trabalho permitem obter clones multicópia, foi avaliado o efeito do número de cópias na produção de ambas as proteínas (GFP e amilase). Como mostrado na figura 64, existe uma correlação linear entre a produção de GFP (fluorescência medida por citometria de fluxo) e o número de cópias do vetor para os clones transformados com o vetor pKld-GFP ( $R^2 = 0.8757$ ).

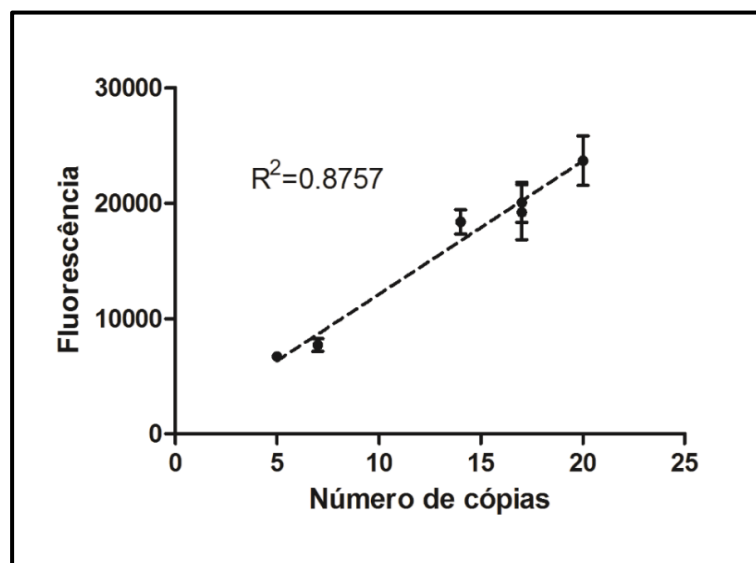
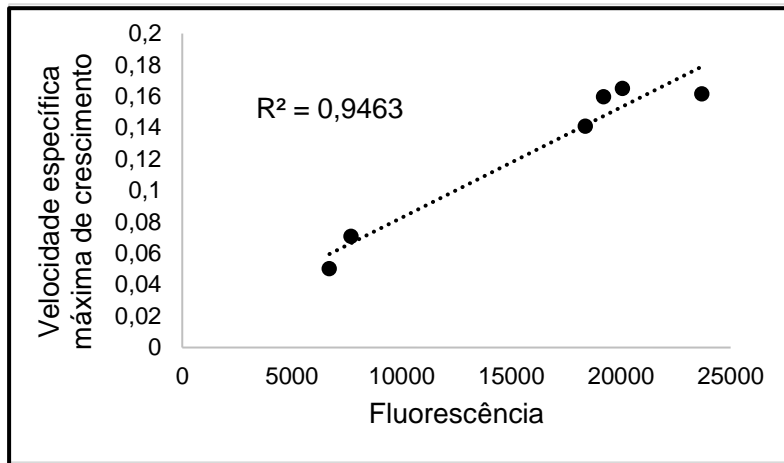


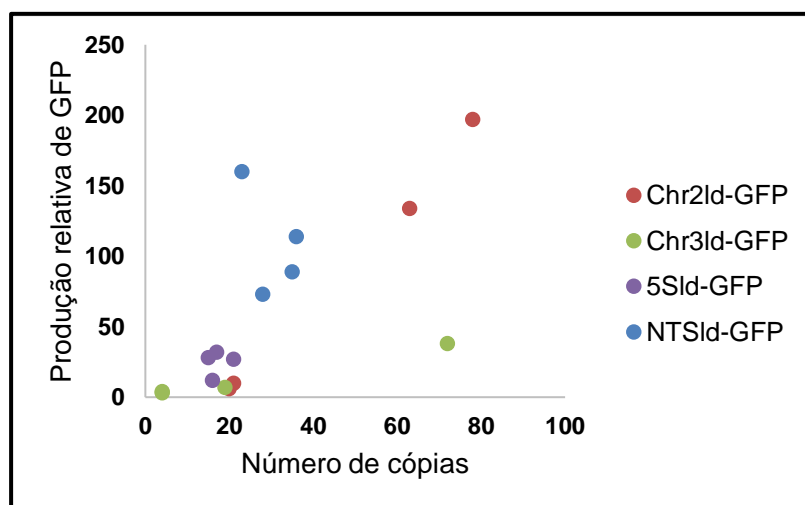
Figura 64. Correlação entre a fluorescência e o número de cópias para os clones transformados com o vetor pKld-GFP. As barras de erro representam o erro padrão da média para  $n=3$ .

É esperado que a partir de um determinado número de cópias a produção da proteína heteróloga seja prejudicial para a célula devido à limitação de outros processos celulares pela competição de substratos ou precursores. Todavia, no caso dos clones analisados transformados com o vetor pKId-GFP, o crescimento celular não foi afetado e, inclusive, foi observada uma correlação linear entre a velocidade específica máxima de crescimento e a fluorescência ( $R^2 = 0,9463$ ) como observado na figura 65.



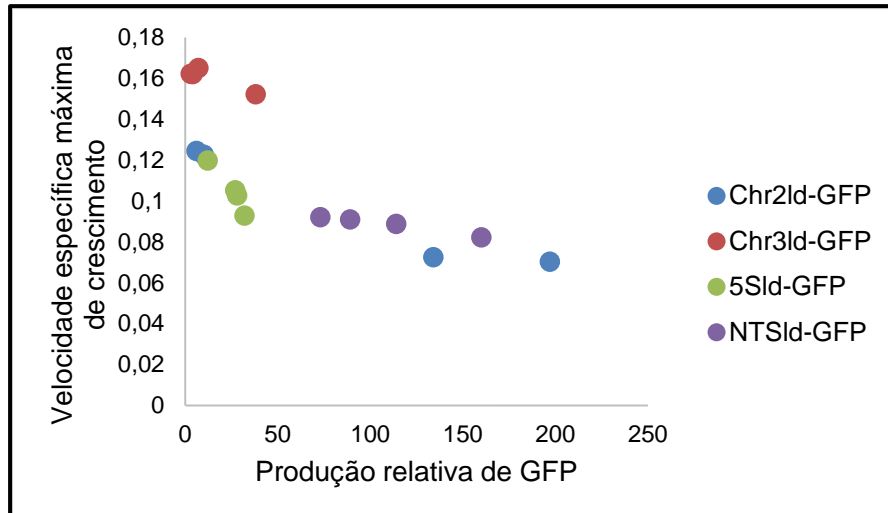
**Figura 65.** Correlação entre a velocidade específica máxima de crescimento e a fluorescência para os clones transformados com o vetor pKId-GFP.

Quando foram usados os cassetes de expressão contendo as sequências repetidas, não foi observada correlação entre a produção de GFP e o número de cópias do cassete (figura 66). Porém, comparando os clones transformados com o mesmo cassete não foi observado um efeito negativo na produção com o aumento no número de cópias.



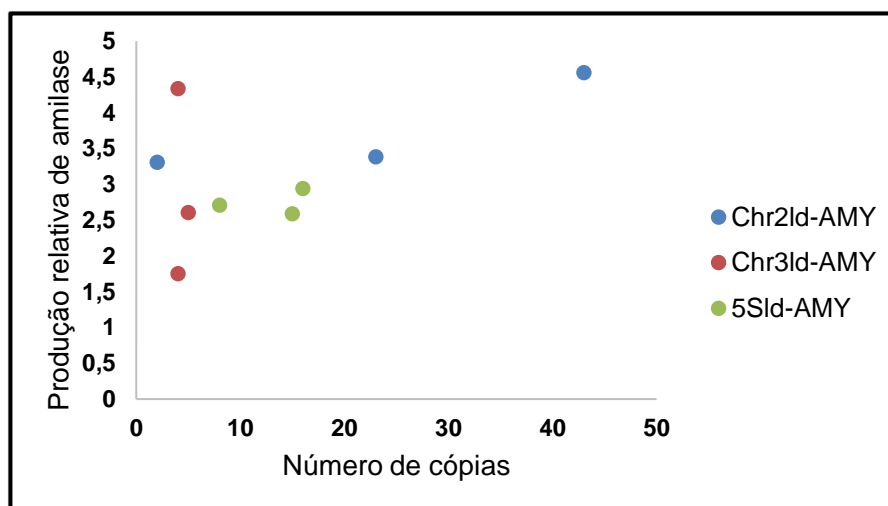
**Figura 66.** Correlação entre a fluorescência e o número de cópias para os clones transformados com os cassetes contendo as sequências repetidas e o gene EGFP.

O aumento na produção de GFP com os clones transformados com as sequências repetidas teve um efeito negativo no crescimento da levedura. Os clones com maiores valores de produção tiveram uma redução na velocidade de crescimento (figura 67).



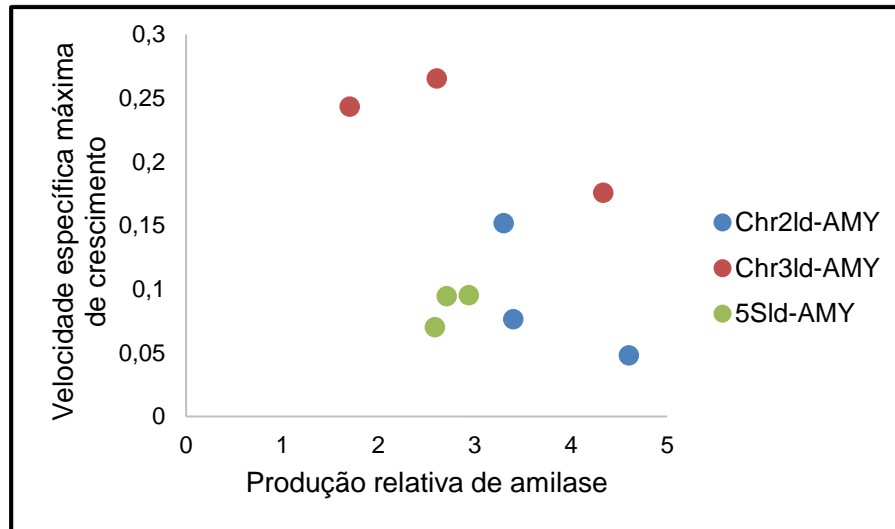
**Figura 67.** Correlação entre a velocidade específica máxima de crescimento e a produção de GFP para os clones transformados com os cassetes com sequências repetidas.

Os clones transformados com os cassetes contendo as sequências repetidas e o gene *AMY1* não apresentaram correlação linear entre a produção de amilase e o número de cópias (figura 68). Contudo, o aumento do número de cópias também não levou a uma diminuição da produção da proteína.



**Figura 68.** Correlação entre a atividade amilolítica e o número de cópias para os clones transformados com os cassetes contendo as sequências repetidas e o gene *AMY1*.

O aumento na produção de amilase teve um efeito negativo no crescimento dos clones transformados com os cassetes Chr2ld-AMY e 5Sld-AMY. Esse efeito não foi observado nos clones transformados com o cassete Chr3ld-AMY (figura 69).



**Figura 69.** Correlação entre a velocidade específica máxima de crescimento e a produção de amilase para os clones transformados com os cassetes com sequências repetidas.

A relação positiva entre a produção de GFP e o número de cópias está em concordância com um estudo anterior onde ao aumentar até oito o número de cópias do antígeno de superfície da hepatite B (HBsAg) em *K. phaffii*, foi observada uma correlação linear com a concentração de mRNA e da proteína traduzida (Vassileva *et al.*, 2001). Igualmente, uma correlação direta foi observada entre o número de cópias do gene e a secreção de miniproinsulina (MPI) em *K. phaffii* (Mansur *et al.*, 2005).

Contudo, a relação entre o número de cópias do gene e a produção da proteína não é sempre linear e, em alguns casos, pode ser até prejudicial, especialmente quando se trata de proteínas secretadas devido a um estresse adicional na via de secreção (Aw e Polizzi, 2013). Em outro estudo com EGFP, foi observado aumento da proteína secretada ao aumentar o número de cópias do gene sequencialmente até três cópias, mas a produção foi diminuída com seis cópias do gene (Liu *et al.*, 2014).

Embora o aumento do número de cópias de ambos os genes utilizados neste trabalho não tenha mostrado um efeito negativo na produção da proteína, o aumento da produção foi muito mais notável para GFP do que para a amilase. Setenta e oito cópias do cassete Chr2ld-GFP aumentaram a produção de GFP em até 197 vezes, enquanto que 43 cópias do cassete

Chr2ld-AMY resultaram em um aumento da produção de amilase em apenas 4,6 vezes. Além disso, o aumento na produção do clone com 23 cópias do cassete Chr2ld-AMY não foi muito maior do que o observado com o clone com duas cópias integradas do mesmo cassete. É importante notar que o tamanho da amilase é quase o dobro do tamanho da GFP, o que significa uma maior necessidade de aminoácidos e precursores para sintetizar a amilase.

O fato de ter identificado atividade amilolítica no extrato proteico intracelular confirmou que parte da proteína está ficando retida dentro da célula, o que também poderia explicar o aumento mais sutil na produção de amilase em comparação a GFP. Este resultado está em concordância com o observado em trabalhos anteriores onde o uso de clones multicópia para produzir uma proteína secretada apresentou baixa produtividade ou até mesmo a diminuição da produção. Quando clones multicópia foram usados para produzir superóxido dismutase humana (hSOD) intracelularmente em *K. phaffii* e albumina do soro humano (HSA) secretada em *K. pastoris*, foi observada uma diferença na correlação entre número de cópias e produção para as proteína não-secretada e secretada (Marx *et al.*, 2009). A produção de hSOD mostrou correlação linear com o número de cópias do gene, enquanto que a produção de HSA aumentou até aproximadamente sete cópias e diminuiu com números de cópias maiores.

Linhagens de *K. phaffii* secretando tripsinogênio humano sob controle do promotor *AOX1*, apresentaram uma diminuição da produção da proteína quando três cópias do cassete de expressão foram integradas, sendo que uma grande parte da proteína foi retida na fração insolúvel do retículo endoplasmático (Hohenblum *et al.*, 2004). Estes estudos mostraram que gargalos na via de secreção são, em grande parte, responsáveis pela baixa produtividade de alguns clones multicópia havendo, no entanto, uma variação no número de cópias integradas necessária para atingir a saturação de produção que varia para cada proteína (Aw e Polizzi, 2013). Fica evidente, pois, que a otimização da via de secreção de *K. phaffii* é uma abordagem necessária para melhorar a produção de proteínas secretadas.

## 6. CONCLUSÕES E PERSPECTIVAS

- É possível usar o alelo defectivo *leu2-d* de *S. cerevisiae* como marca de seleção em vetores de expressão para *K. phaffii* e a presença dessa marca favorece a integração de múltiplas cópias do vetor, ou cassete de expressão, no genoma da levedura.
- O tamanho das sequências repetidas avaliadas neste trabalho não é suficientemente grande para promover uma integração específica no *locus* desejado, porém, a presença dessas sequências no vetor de expressão leva a um aumento no número de cópias integradas, devido, provavelmente, a uma maior quantidade de *loci* disponíveis para recombinação não homóloga.
- O *locus* 5S mostrou ser uma região não-estável e, apesar de estar em 21 cópias no genoma, apenas uma *locus* foi alvo de integração.
- O uso da marca *leu2-d* junto com sequências repetidas como alvo de integração pode ser considerado como uma primeira etapa na construção de linhagens com maiores produtividades. Com esta abordagem é possível obter um painel de clones com diferentes números de cópias que podem ser avaliados para a aplicação desejada.
- O aumento da produção da proteína heteróloga nos clones multicópia foi mais acentuado no caso da proteína intracelular (GFP) do que no caso da proteína secretada (amilase).
- Não é possível definir qual dos sistemas aqui testados é o melhor para a expressão heteróloga. Para cada proteína existe um sistema que proporciona a melhor relação entre número de cópias e produção. Além disso, é mais importante procurar pelo número de cópias ótimo do que o máximo número de cópias.
- É necessário melhorar a via de secreção de *K. phaffii* para aumentar a produção de proteínas secretadas.
- Uma vez tendo o clone com o número de cópias que otimiza a produção da proteína de interesse, a seguinte etapa é otimizar o processo de produção em larga escala com condições de crescimento controladas.



## 7. REFERÊNCIAS

- Ahmad, M.; Hirz, M.; Pichler, H.; e Schwab, H., (2014). Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology* 98(12), 5301–5317. doi:10.1007/s00253-014-5732-5
- Araújo, J. D. A., (2008). Produção de quimosina B de *Bos taurus* em *Pichia pastoris*. Tese de doutorado apresentada ao programa de Biologia Celular do Instituto de Ciências Biológicas. Universidade de Brasília.
- Arruda, A.; Reis, V. C. B.; Batista, V. D. F.; Daher, B. S.; Piva, L. C.; de Marco, J. L.; de Moraes, L. M. P.; e Torres, F. A. G., (2016). A constitutive expression system for *Pichia pastoris* based on the *PGK1* promoter. *Biotechnology Letters* 38(3), 509–517. doi:10.1007/s10529-015-2002-2
- Aw, R.; e Polizzi, K. M., (2016). Liquid PTV: a faster and cheaper alternative for generating multi-copy clones in *Pichia pastoris*. *Microbial Cell Factories* 15(1), 29–39. doi:10.1186/s12934-016-0432-8
- Aw, R.; e Polizzi, K. M., (2013). Can too many copies spoil the broth? *Microbial Cell Factories* 12, 128–136. doi:10.1186/1475-2859-12-128
- Benson, G., (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27(2), 573–580.
- Betancur, M. O.; Reis, V. C. B.; Nicola, A. M.; De Marco, J. L.; de Moraes, L. M. P.; e Torres, F. A. G., (2017). Multicopy plasmid integration in *Komagataella phaffii* mediated by a defective auxotrophic marker. *Microbial Cell Factories* 16(1), 99. doi:10.1186/s12934-017-0715-8
- Camattari, A.; Goh, A.; Yip, L. Y.; Tan, A. H. M.; Ng, S. W.; Tran, A.; Liu, G.; Liachko, I.; Dunham, M. J.; e Rancati, G., (2016). Characterization of a panARS-based episomal vector in the methylotrophic yeast *Pichia pastoris* for recombinant protein production and synthetic biology applications. *Microbial Cell Factories* 15(1), 139. doi:10.1186/s12934-016-0540-5
- Çelik, E.; e Çalık, P., (2012). Production of recombinant proteins by yeast cells. *Biotechnology Advances* 30(5), 1108–1118. doi:10.1016/j.biotechadv.2011.09.011
- Cereghino, J. L.; e Cregg, J. M., (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews* 24(1), 45–66. doi:10.1111/j.1574-6976.2000.tb00532.x
- Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W.; e Prasher, D., (1994). Green fluorescent protein as a marker gene expression. *Science* 263, 802–805. doi:10.1126/science.8303295

- Chen, H.-L.; Yen, C.-C.; Tsai, T.-C.; Yu, C.-H.; Liou, Y.-J.; Lai, Y.-W.; Wang, M.-L.; e Chen, C.-M., (2006). Production and characterization of human extracellular superoxide dismutase in the methylotrophic yeast *Pichia pastoris*. *Journal of Agricultural and Food Chemistry* 54(21), 8041–8047. doi:10.1021/jf061379x
- Ciofalo, V.; Barton, N.; Kreps, J.; Coats, I.; e Shanahan, D., (2006). Safety evaluation of a lipase enzyme preparation, expressed in *Pichia pastoris*, intended for use in the degumming of edible vegetable oil. *Regulatory Toxicology and Pharmacology* 45(1), 1–8. doi:10.1016/j.yrtph.2006.02.001
- Clare, J. J.; Rayment, F. B.; Ballantine, S. P.; Sreekrishna, K.; e Romanos, M. A., (1991a). High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Bio/Technology* 9, 455–460.
- Clare, J. J.; Romanos, M. a.; Rayment, F. B.; Rowedder, J. E.; Smith, M. a.; Payne, M. M.; Sreekrishna, K.; e Henwood, C. a., (1991b). Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. *Gene* 105(2), 205–212. doi:10.1016/0378-1119(91)90152-2
- Cormack, B. P.; Valdivia, R. H.; e Falkow, S., (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173, 33–38. doi:10.1016/0378-1119(95)00685-0
- Cos, O.; Serrano, A.; Montesinos, J. L.; Ferrer, P.; Cregg, J. M.; e Valero, F., (2005). Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures. *Journal of Biotechnology* 117(1), 321–335. doi:10.1016/j.jbiotec.2004.12.010
- Costantini, L. M.; Fossati, M.; Francolini, M.; e Snapp, E. L., (2013). Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions. *Traffic* 13(5), 643–649. doi:10.1111/j.1600-0854.2012.01336.x.
- Cregg, J. M.; Barringer, K. J.; Hessler, A. Y.; e Madden, K. R., (1985). *Pichia pastoris* as a Host System for Transformations. *Molecular and Cellular Biology* 5(12), 3376–3385. doi:10.1128/MCB.5.12.3376.Updated
- Cregg, J. M.; Cereghino, J. L.; Shi, J.; e Higgins, D. R., (2000). Recombinant Protein Expression in *Pichia pastoris*. *Molecular Biotechnology* 16(1), 23–52. doi:10.1385/MB:16:1:23
- Cregg, J. M.; Tschopp, J. F.; Stillman, C.; Siegel, R.; Akong, M.; Craig, W. S.; Buckholz, R. G.; Madden, K. R.; Kellaris, P. a; Davis, G. R.; Smiley, B. L.; Cruze, J.; Torregrossa, R.; Velicelebi, G.; e Thill, G. P., (1987). High-level expression and efficient assembly of hepatitis B surface antigen in the methylotrophic yeast, *Pichia pastoris*. *Bio/Technology* 5(5), 479–485. doi:10.1038/nbt0587-479
- De Almeida, J. R. M.; Moraes, L. M.; e Torres, F. A., (2005). Molecular characterization of the 3-phosphoglycerate kinase gene (*PGK1*) from the methylotrophic yeast *Pichia pastoris*. *Yeast* 22, 725–737. doi:10.1002/yea.1243

- De Schutter, K.; Lin, Y.-C.; Tiels, P.; Van Hecke, A.; Glinka, S.; Weber-Lehmann, J.; Rouzé, P.; Van de Peer, Y.; e Callewaert, N., (2009). Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nature Biotechnology* 27(6), 561–566. doi:10.1038/nbt.1544
- Demeke, M. M.; Foulquié-Moreno, M. R.; Dumortier, F.; e Thevelein, J. M., (2015). Rapid Evolution of Recombinant *Saccharomyces cerevisiae* for Xylose Fermentation through Formation of Extra-chromosomal Circular DNA. *PLoS Genetics* 11(3), 1–21. doi:10.1371/journal.pgen.1005010
- Dorsey, M.; Peterson, C.; Bray, K.; e Paquin, C. E., (1992). Spontaneous amplification of the *ADH4* gene in *Saccharomyces cerevisiae*. *Genetics* 132(4), 943–950.
- Dröge, M.; Pühler, A.; e Selbitschka, W., (1998). Horizontal gene transfer as a biosafety issue: a natural phenomenon of public concern. *Journal of Biotechnology* 64(1), 75–90. doi:10.1016/S0168-1656(98)00105-9
- Du, M.; Battles, M. B.; e Nett, J. H., (2012). A color-based stable multi-copy integrant selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as auxotrophic markers. *Bioengineered Bugs* 3(1), 32–37. doi:10.4161/bbug.3.1.17936
- Erhart, E.; e Hollenberg, C. P., (1983). The presence of a defective *LEU2* gene on 2 mu DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number. *Journal of Bacteriology* 156(2), 625–635.
- FDA, (1993). Food and Drug Administration. Food Additives Permitted in Feed and Drinking Water of Animals. *Pichia pastoris* Dried Yeast.
- Flagfeldt, D. B.; Siewers, V.; Huang, L.; e Nielsen, J., (2009). Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* 26, 545–551. doi:10.1002/yea
- Fuwa, H., (1954). A new method for microdetermination of amylase activity by the use of amylose as the substrate. *The Journal of Biochemistry* 41(5). doi:10.1093/oxfordjournals.jbchem.a126476
- Galdino, A. S., (2008). Clonagem e expressão de uma  $\alpha$ -amilase de *Cryptococcus flavus* e sua aplicação na degradação do amido 139.
- Galdino, A. S.; Silva, R. N.; Lottermann, M. T.; Álvares, A. C. M.; Moraes, L. M. P. De; Torres, F. A. G.; Freitas, S. M. De; Ulhoa, C. J.; Alves; Freitas, S. M. De; e Ulhoa, C. J., (2011). Biochemical and Structural Characterization of Amy1: An Alpha-Amylase from *Cryptococcus flavus* Expressed in *Saccharomyces cerevisiae*. *Enzyme Research* 2011, 1–7. doi:10.4061/2011/157294
- Galdino, A. S.; Ulhoa, C. J.; Moraes, L. M. P.; Prates, M. V.; Bloch, C.; e Torres, F. a G., (2008). Cloning, molecular characterization and heterologous expression of *AMY1*, an  $\alpha$ -amylase gene from *Cryptococcus flavus*. *FEMS Microbiology Letters* 280(2), 189–194. doi:10.1111/j.1574-6968.2007.01059.x

- Gasser, B.; Prielhofer, R.; Marx, H.; Maurer, M.; Nocon, J.; Steiger, M.; Puxbaum, V.; Sauer, M.; e Mattanovich, D., (2013). *Pichia pastoris*: protein production host and model organism for biomedical research. *Future Microbiology* 8(2), 191–208.
- Guilliermond, A., (1920). *Zygosaccharomyces pastori*, nouvelle espèce de levures copulation hétérogamique. *Bulletin de la Société Mycologique de France* 36, 203–211.
- Gupta, R.; Gigras, P.; Mohapatra, H.; Goswami, V. K.; e Chauhan, B., (2003). Microbial  $\alpha$ -amylases: a biotechnological perspective. *Process Biochemistry* 38(11), 1599–1616. doi:10.1016/S0032-9592(03)00053-0
- Higgins, D. R.; e Cregg, J. M., (1998). *Pichia* protocols. Humana Press.
- Hohenblum, H.; Gasser, B.; Maurer, M.; Borth, N.; e Mattanovich, D., (2004). Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnology and Bioengineering* 85(4), 367–375. doi:10.1002/bit.10904
- Hoylaerts, M.; Weyens, A.; Bollen, A.; Harford, N.; e Cabezón, T., (1986). High-level production and isolation of human recombinant alpha 1-proteinase inhibitor in yeast. *FEBS letters* 204(1), 83–87. doi:10.1016/0014-5793(86)81391-6
- Iefuji, H.; Chino, M.; Kato, M.; e Imura, Y., (1996). Raw-starch-digesting and thermostable  $\alpha$ -amilase from the yeast *Cryptococcus* sp. S-2: purification, characterization, cloning and sequencing. *Biochemical Journal* 318, 989–996.
- Janner, C. R.; Brito, A. L. P.; Moraes, L. M. P.; Reis, V. C.; e Torres, F. A., (2013). pPCV, a versatile vector for cloning PCR products. *SpringerPlus* 2, 441–443. doi:10.1186/2193-1801-2-441
- Kang, Z.; Huang, H.; Zhang, Y.; Du, G.; e Chen, J., (2017). Recent advances of molecular toolbox construction expand *Pichia pastoris* in synthetic biology applications. *World Journal of Microbiology and Biotechnology* 33, 19–27. doi:10.1007/s11274-016-2185-2
- Kazemi Seresht, A.; Nørgaard, P.; Palmqvist, E. A.; Andersen, A. S.; e Olsson, L., (2013). Modulating heterologous protein production in yeast: the applicability of truncated auxotrophic markers. *Applied Microbiology and Biotechnology* 97, 3939–3948. doi:10.1007/s00253-012-4263-1
- Kobayashi, T.; Heck, D. J.; Nomura, M.; e Horiuchi, T., (1998). Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes & Development* 12(24), 3821–3830. doi:10.1101/gad.12.24.3821
- Kobayashi, T.; Nomura, M.; e Horiuchi, T., (2001). Identification of DNA cis elements essential for expansion of ribosomal DNA repeats in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 21(1), 136–147. doi:10.1128/MCB.21.1.136
- Kuberl, A.; Schneider, J.; Thallinger, G. G.; Anderl, I.; Wibberg, D.; Hajek, T.; Jaenicke, S.; Brinkrolf, K.; Goesmann, A.; Szczepanowski, R.; Pühler, A.; Schwab, H.; Glieder, A.; e

- Pichler, H., (2011). High-quality genome sequence of *Pichia pastoris* CBS7435. *Journal of Biotechnology* 154, 312–320. doi:10.1016/j.jbiotec.2011.04.014
- Kurtzman, C. P., (2009). Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *Journal of Industrial Microbiology and Biotechnology* 36, 1435–1438. doi:10.1007/s10295-009-0638-4
- Lee, F. W.; e Da Silva, N. A., (1997). Improved efficiency and stability of multiple cloned gene insertions at the delta sequences of *Saccharomyces cerevisiae*. *Applied microbiology and biotechnology* 48(3), 339–345.
- Li, Y. T.; Li, M. T.; Fu, C. H.; Zhou, P. P.; Liu, J. M.; e Yu, L. J., (2009). Improvement of arachidonic acid and eicosapentaenoic acid production by increasing the copy number of the genes encoding fatty acid desaturase and elongase into *Pichia pastoris*. *Biotechnology Letters* 31(7), 1011–1017. doi:10.1007/s10529-009-9970-z
- Lim, H.-K.; Kim, K.-Y.; Lee, K.-J.; Park, D.-H.; Chung, S.-I.; e Jung, K.-H., (2000). Genetic stability of the integrated structural gene of guamerin in recombinant *Pichia pastoris*. *Journal of Microbiology and Biotechnology* 10(4), 470–475.
- Liu, H.; Qin, Y.; Huang, Y.; Chen, Y.; Cong, P.; e He, Z., (2014). Direct evaluation of the effect of gene dosage on secretion of protein from yeast *Pichia pastoris* by expressing EGFP. *J.Microbiol. Biotechnol.* 24(2), 144–151. doi:10.4014/jmb.1308.08010
- Lopes, T. S.; Hakkaart, G.-J. A. J.; Koerts, B. L.; Raué, H. A.; e Planta, R. J., (1991). Mechanism of high-copy-number integration of pMIRY-type vectors into the ribosomal DNA of *Saccharomyces cerevisiae*. *Gene* 105(1), 83–90. doi:10.1016/0378-1119(91)90516-E
- Love, K. R.; Politano, T. J.; Panagiotou, V.; Jiang, B.; Stadheim, T. a.; e Love, J. C., (2012). Systematic single-cell analysis of *Pichia pastoris* reveals secretory capacity limits productivity. *PLoS ONE* 7(6), 1–11. doi:10.1371/journal.pone.0037915
- Love, K. R.; Shah, K. a.; Whittaker, C. a.; Wu, J.; Bartlett, M. C.; Ma, D.; Leeson, R. L.; Priest, M.; Borowsky, J.; Young, S. K.; e Love, J. C., (2016). Comparative genomics and transcriptomics of *Pichia pastoris*. *BMC Genomics* 17, 550. doi:10.1186/s12864-016-2876-y
- Macauley-Patrick, S.; Fazenda, M. L.; McNeil, B.; e Harvey, L. M., (2005). Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 22(4), 249–270. doi:10.1002/yea.1208
- Mansur, M.; Cabello, C.; Hernández, L.; País, J.; Varas, L.; Valdés, J.; Terrero, Y.; Hidalgo, A.; Plana, L.; Besada, V.; García, L.; Lamazares, E.; Castellanos, L.; e Martínez, E., (2005). Multiple gene copy number enhances insulin precursor secretion in the yeast *Pichia pastoris*. *Biotechnology Letters* 27(5), 339–345. doi:10.1007/s10529-005-1007-7

- Marx, H.; Mecklenbräuker, A.; Gasser, B.; Sauer, M.; e Mattanovich, D., (2009). Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. *FEMS Yeast Research* 9(8), 1260–1270. doi:10.1111/j.1567-1364.2009.00561.x
- Mattanovich, D.; Gasser, B.; Hohenblum, H.; e Sauer, M., (2004). Stress in recombinant protein producing yeasts. *Journal of Biotechnology* 113(1-3), 121–135. doi:10.1016/j.jbiotec.2004.04.035
- Mattanovich, D.; Graf, A.; Stadlmann, J.; Dragosits, M.; Redl, A.; Maurer, M.; Kleinheinz, M.; Sauer, M.; Altmann, F.; e Gasser, B., (2009). Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microbial Cell Factories* 8(1), 29–41. doi:10.1186/1475-2859-8-29
- Montesino, R.; García, R.; Quintero, O.; e Cremata, J. A., (1998). Variation in N-Linked Oligosaccharide Structures on Heterologous Proteins Secreted by the Methylophilic Yeast *Pichia pastoris*. *Protein Expression and Purification* 14(2), 197–207. doi:10.1006/pep.1998.0933
- Moon, H. K.; Lee, D. W.; Sim, G. H.; Kim, H.-J.; Hwang, J. Y.; Kwon, M.-G.; Kang, B.-K.; Kim, J. M.; e Kang, H. A., (2016). A new set of rDNA-NTS-based multiple integrative cassettes for the development of antibiotic-marker-free recombinant yeasts. *Journal of Biotechnology* 233, 190–199. doi:10.1016/j.jbiotec.2016.07.006
- Näätsaari, L.; Mistlberger, B.; Ruth, C.; Hajek, T.; Hartner, F. S.; e Glieder, A., (2012). Deletion of the *Pichia pastoris* *KU70* homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS ONE* 7(6), e39720. doi:10.1371/journal.pone.0039720
- Nett, J. H.; e Gerngross, T. U., (2003). Cloning and disruption of the PpURA5 gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*. *Yeast* 20(15), 1279–1290. doi:10.1002/yea.1049
- Nieto, A.; Prieto, J. a; e Sanz, P., (1999). Stable high-copy-number integration of *Aspergillus oryzae* alpha-amylase cDNA in an industrial baker's yeast strain. *Biotechnology Progress* 15, 459–466.
- Ohi, H.; Okazaki, N.; Uno, S.; Miura, M.; e Hiramatsu, R., (1998). Chromosomal DNA patterns and gene stability of *Pichia pastoris*. *Yeast* 14(10), 895–903. doi:10.1002/(SICI)1097-0061(199807)14:10<895::AID-YEA288>3.0.CO;2-9
- Paquin, C. E.; Dorsey, M.; Crable, S.; Sprinkel, K.; Sondej, M.; e Williamson, V. M., (1992). A spontaneous chromosomal amplification of the *ADH2* gene in *Saccharomyces cerevisiae*. *Genetics* 130(2), 263–271.
- Pignede, G.; Wang, H.-J.; Fudalej, F.; Seman, M.; Gaillardin, C.; e Nicaud, J.-M., (2000). Autocloning and Amplification of *LIP2* in *Yarrowia lipolytica*. *Applied and Environmental Microbiology* 66(8), 3283–3289. doi:10.1128/AEM.66.8.3283-3289.2000

- Piva, L. C.; Betancur, M. O.; Reis, V. C. B.; De Marco, J. L.; Moraes, L. M. P.; e Torres, F. A. G., (2017). Molecular strategies to increase the levels of heterologous transcripts in *Komagataella phaffii* for protein production. *Bioengineered* 0, 1–5. doi:10.1080/21655979.2017.1296613
- Ro, D.-K.; Ouellet, M.; Paradise, E. M.; Burd, H.; Eng, D.; Paddon, C. J.; Newman, J. D.; e Keasling, J. D., (2008). Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinin acid. *BMC Biotechnology* 8, 83–96. doi:10.1186/1472-6750-8-83
- Sambrook, J.; Fritsch, E. F.; e Maniatis, T., (1989). *Molecular cloning: a laboratory manual*, 2° ed. New York.
- Servienė, E.; e Melvydas, V., (2001). Effect of the defective leucine gene *leu2-d*, on the properties of recombinant plasmid in yeast *Saccharomyces cerevisiae*. *Biologija* 4, 30–33.
- Souza, P. M.; e Magalhães, P. D. O., (2010). Application of microbial  $\alpha$ -amylase in industry - a review. *Brazilian Journal of Microbiology* 41(4), 850–861. doi:10.1590/S1517-83822010000400004
- Sreekrishna, K.; e Kropp, K. E., (1996). *Pichia pastoris*, em: K. Wolf, *Nonconventional Yeasts in Biotechnology*. Springer, p. 203–253.
- Sturmberger, L.; Chappell, T.; Geier, M.; Krainer, F.; Day, K. J.; Vide, U.; Trstenjak, S.; Schiefer, A.; Richardson, T.; Soriaga, L.; Darnhofer, B.; Birner-Gruenberger, R.; Glick, B. S.; Tolstorukov, I.; Cregg, J.; Madden, K.; e Glieder, A., (2016). Refined *Pichia pastoris* reference genome sequence. *Journal of Biotechnology*. doi:10.1016/j.jbiotec.2016.04.023
- Sunga, A. J.; Tolstorukov, I.; e Cregg, J. M., (2008). Posttransformational vector amplification in the yeast *Pichia pastoris*. *FEMS Yeast Research* 8, 870–876. doi:10.1111/j.1567-1364.2008.00410.x
- Thoma, J. A.; e French, D., (1960). The Starch-Iodine-Iodide Interaction. Part I. Spectrophotometric Investigations. *Journal of the American Chemical Society* 82(16), 4144–4147. doi:10.1021/ja01501a004
- Thor, D.; Xiong, S.; Orazem, C. C.; Kwan, A.-C.; Cregg, J. M.; Lin-Cereghino, J.; e Lin-Cereghino, G. P., (2005). Cloning and characterization of the *Pichia pastoris* *MET2* gene as a selectable marker. *FEMS Yeast Research* 5(10), 935–942. doi:10.1016/j.femsyr.2005.03.009
- Tran, A.-M.; Nguyen, T.-T.; Nguyen, C.-T.; Huynh-Thi, X.-M.; Nguyen, C.-T.; Trinh, M.-T.; Tran, L.-T.; Cartwright, S. P.; Bill, R. M.; e Tran-Van, H., (2017). *Pichia pastoris* versus *Saccharomyces cerevisiae*: a case study on the recombinant production of human granulocyte-macrophage colony-stimulating factor. *BMC Research Notes* 10, 1–8. doi:10.1186/s13104-017-2471-6

- Trimble, R. B.; Atkinson, P. H.; Tschopp, J. F.; Townsend, R. R.; e Maley, F., (1991). Structure of oligosaccharides on *Saccharomyces SUC2* invertase secreted by the methylotrophic yeast *Pichia pastoris*. *Journal of Biological Chemistry* 266(34), 22807–22817.
- Tschopp, J. F.; Sverlow, G.; Kosson, R.; Craig, W.; e Grinna, L., (1987). High-Level Secretion of Glycosylated Invertase in the Methylotrophic Yeast, *Pichia pastoris*. *Bio/Technology* 5(12), 1305–1308. doi:10.1038/nbt1287-1305
- Valli, M.; Tatto, N. E.; Peymann, A.; Gruber, C.; Landes, N.; Ekker, H.; Thallinger, G. G.; Mattanovich, D.; Gasser, B.; e Graf, A. B., (2016). Curation of the genome annotation of *Pichia pastoris* (*Komagataella phaffii*) CBS7435 from gene level to protein function. *FEMS Yeast Research*. doi:10.1093/femsyr/fow051
- Vassileva, A.; Chugh, D.; Swaminathan, S.; e Khanna, N., (2001). Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. *Protein Expression and Purification* 21(1), 71–80. doi:10.1006/prev.2000.1335
- Wanderley, K. J.; Torres, F. a G.; Moraes, L. M. P.; e Ulhoa, C. J., (2004). Biochemical characterization of  $\alpha$ -amylase from the yeast *Cryptococcus flavus*. *FEMS Microbiology Letters* 231(2), 165–169. doi:10.1016/S0378-1097(03)00955-8
- Wang, X.; Wang, Z.; e Da Silva, N. A., (1996). G418 selection and stability of cloned genes integrated at chromosomal delta sequences of *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* 49(1), 45–51. doi:10.1002/(SICI)1097-0290(19960105)49:1<45::AID-BIT6>3.0.CO;2-T
- Waterham, H. R.; Digan, M. E.; Koutz, P. J.; Lair, S. V.; e Cregg, J. M., (1997). Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 186(1), 37–44. doi:10.1016/S0378-1119(96)00675-0
- Weninger, A.; Hatzl, A. M.; Schmid, C.; Vogl, T.; e Glieder, A., (2016). Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *Journal of Biotechnology* 235, 139–149. doi:10.1016/j.jbiotec.2016.03.027
- Zhu, T.; Guo, M.; Sun, C.; Qian, J.; Zhuang, Y.; Chu, J.; e Zhang, S., (2009). A systematical investigation on the genetic stability of multi-copy *Pichia pastoris* strains. *Biotechnology Letters* 31(5), 679–684. doi:10.1007/s10529-009-9917-4
- Zhu, T.; Guo, M.; Tang, Z.; Zhang, M.; Zhuang, Y.; Chu, J.; e Zhang, S., (2009). Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast *Pichia pastoris*. *Journal of Applied Microbiology* 107(3), 954–963. doi:10.1111/j.1365-2672.2009.04279.x



## 8. ANEXOS

### Anexo 1: Sequência sonda PGK1

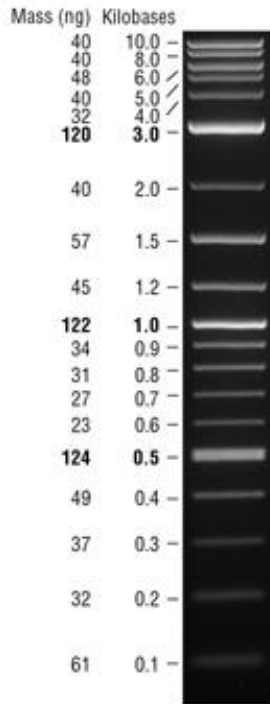
```
gatctgcgag gcaagcatct actaatgttt atttttcgtc caacctaatt gtggtttcaa 60
agcgctatca ggtgggggggt aagaggaatg tgagtggaaa gcgaaaataa ctggcagctg 120
gggtcagatc ccgtgatgcc acctcttgtg gtattttgaa acgctgtgtg cgattggccg 180
cgagaacgga aaggaatata tttactgccc atcgcatttt ggctcaaata aaatcttgag 240
cttttgacaa tagattatat gttctttctt ggaagctctt tcagctaata gtgaagtgtt 300
tcctactaag gatcgcctcc aaacgttcca actacgggag gaggttgcaa agaaaacggg 360
tctctcagcg aattgttctc atccatgagt gagtcctctc cgtcctttcc tcgagcctgg 420
caataaagcc tccttcggag gagctccgtc tagagaataa ttgctgcctt tctgactttc 480
ggactagcgc caaccgagaa ccacaccacc acaccatcac tgtcaccctg catagttcat 540
ccctctctcc ttataaagca tctaataagg tccacaattg tttgccacaa aaatctctta 600
gcatagccca attgattacg aaag 624
```

### Anexo 2: Sequência sonda EGFP

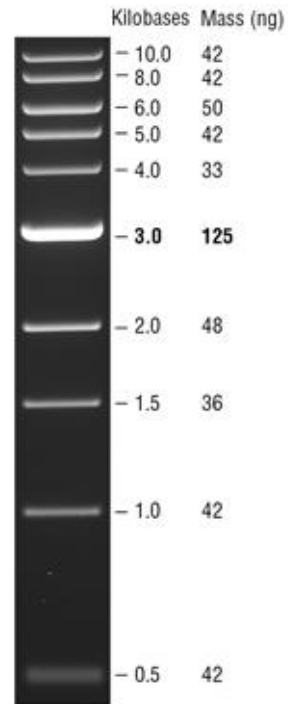
```
ggaattcatg gtgagcaagg gcgaggagct gttcaccggg gtggtgcca tcctggtcga 60
gctggacggc gacgtaaacg gccacaagtt cagcgtgtcc ggcgagggcg agggcgatgc 120
cacctacggc aagctgacct tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg 180
gcccaccctc gtgaccaccc tgacctacgg cgtgcagtgc ttcagccgct accccgacca 240
catgaagcag cacgacttct tcaagtccgc catgcccga ggctacgtcc aggagcgcac 300
catcttcttc aaggacgacg gcaactacaa gaccgcgcc gaggtgaagt tcgagggcga 360
caccctggtg aaccgcatcg agctgaaggg catcgacttc aaggaggacg gcaacatcct 420
ggggcacaag ctggagtaca actacaacag ccacaacgtc tatatcatgg ccgacaagca 480
gaagaacggc atcaaggatg acttcaagat ccgccacaac atcgaggacg gcagcgtgca 540
gctcgccgac cactaccagc agaacacccc catcgccgac ggccccgtgc tgctgcccga 600
caaccactac ctgagcacc agtcgcacct gagcaaagac cccaacgaga agcgcgatca 660
catggtcctg ctggagttcg tgaccgcgcg cgggatcact ctcggcatgg acgagctgta 720
caagtaagcg gccgcc 736
```

### Anexo 3: Padrão dos marcadores de massa molecular usados

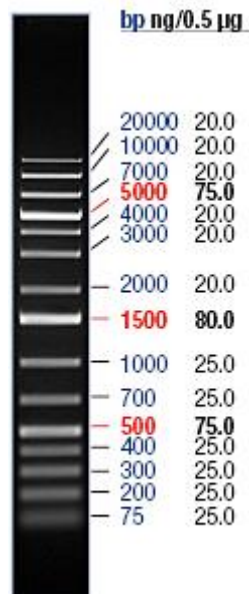
2-log DNA ladder (New England Biolabs)



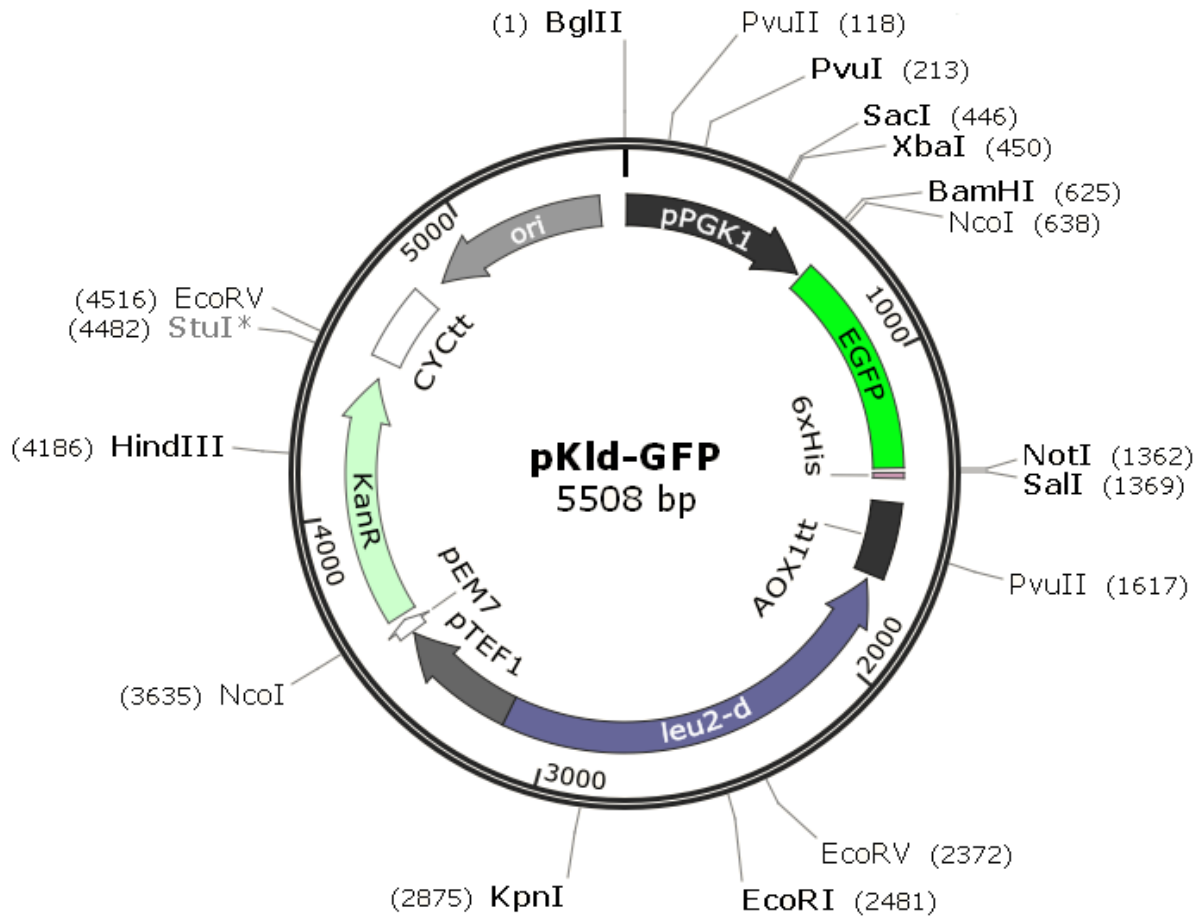
1 kb DNA ladder (New England Biolabs)



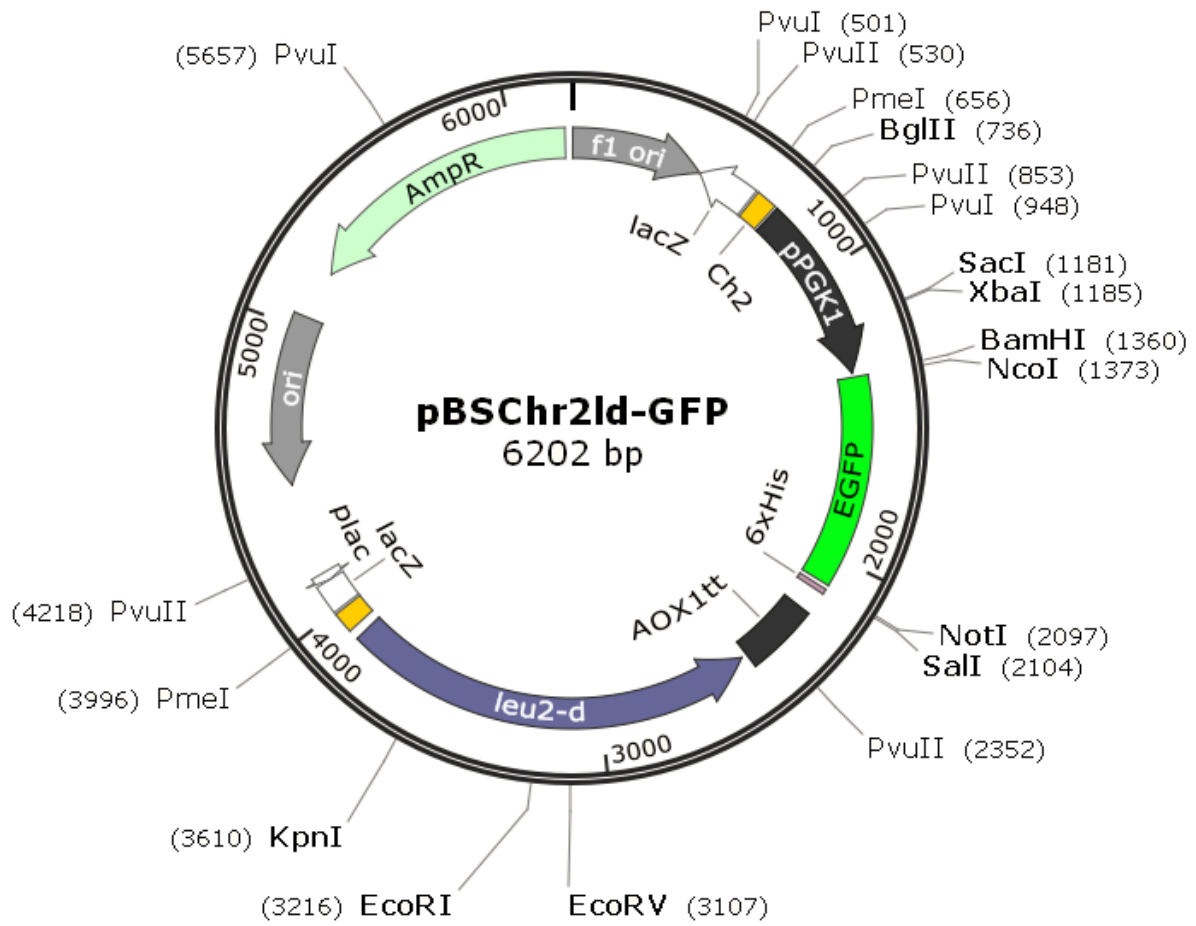
O'GeneRuler 1 kb Plus DNA ladder  
(Thermo Scientific)



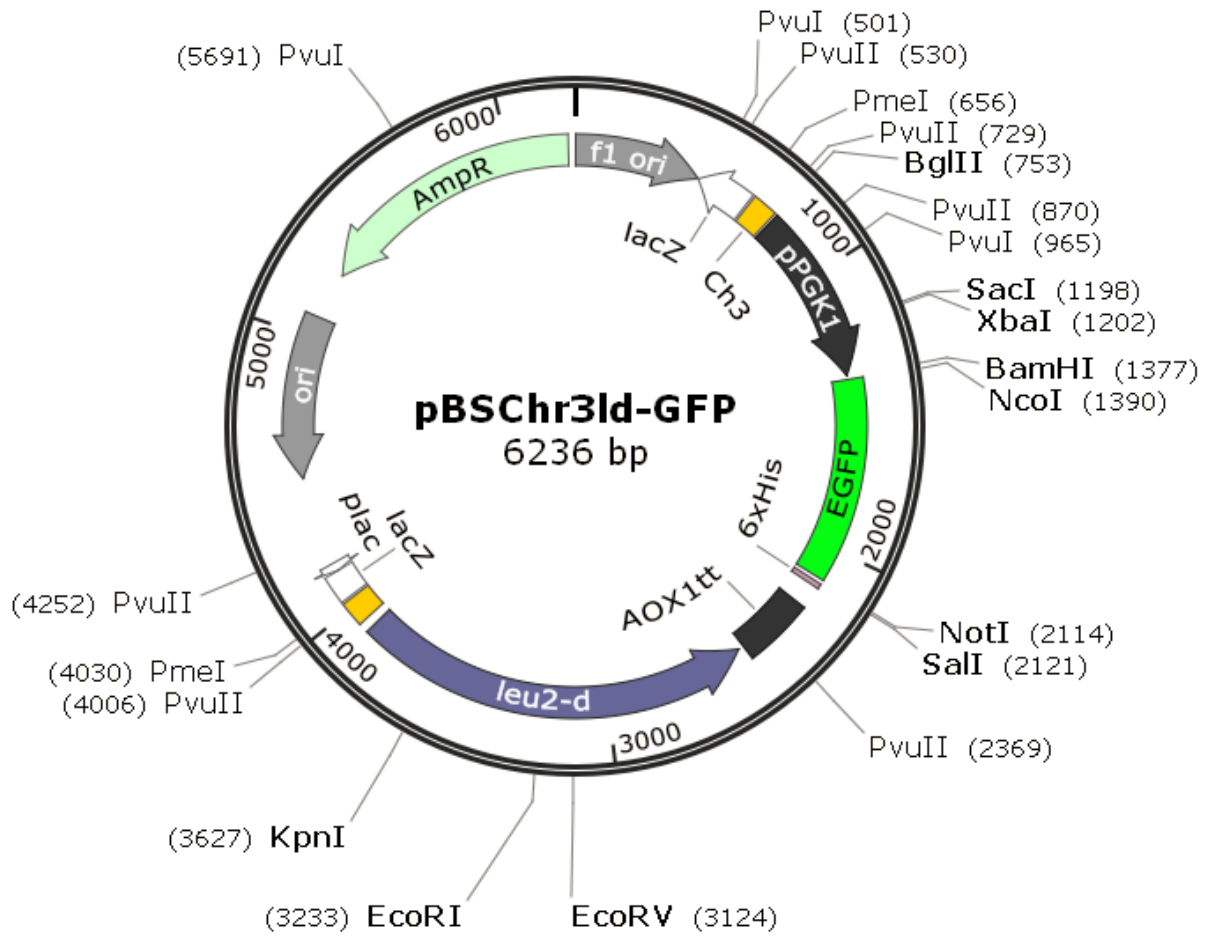
#### Anexo 4: Mapa físico do vetor pKld-GFP



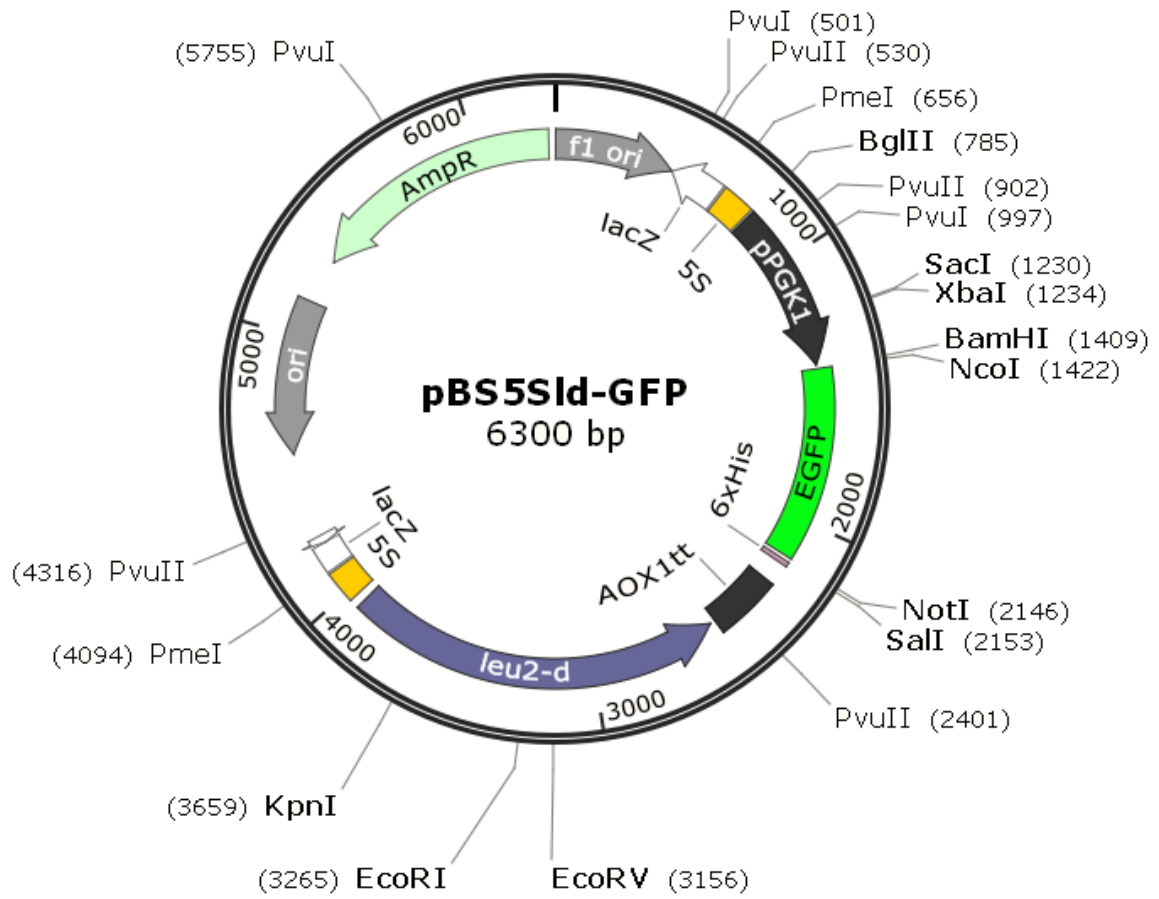
**Anexo 5: Mapa físico do vetor pBSChr2ld-GFP**



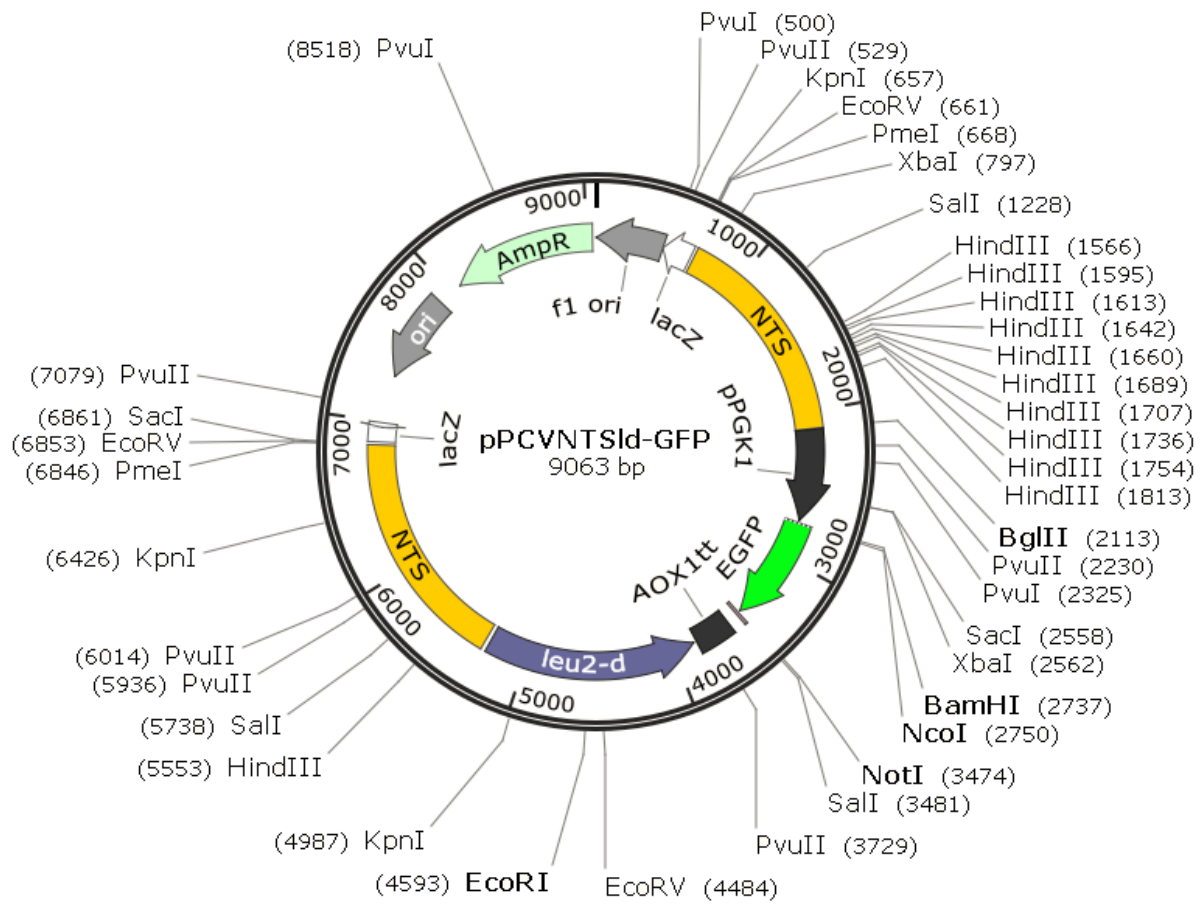
**Anexo 6: Mapa físico do vetor pBSChr3ld-GFP**



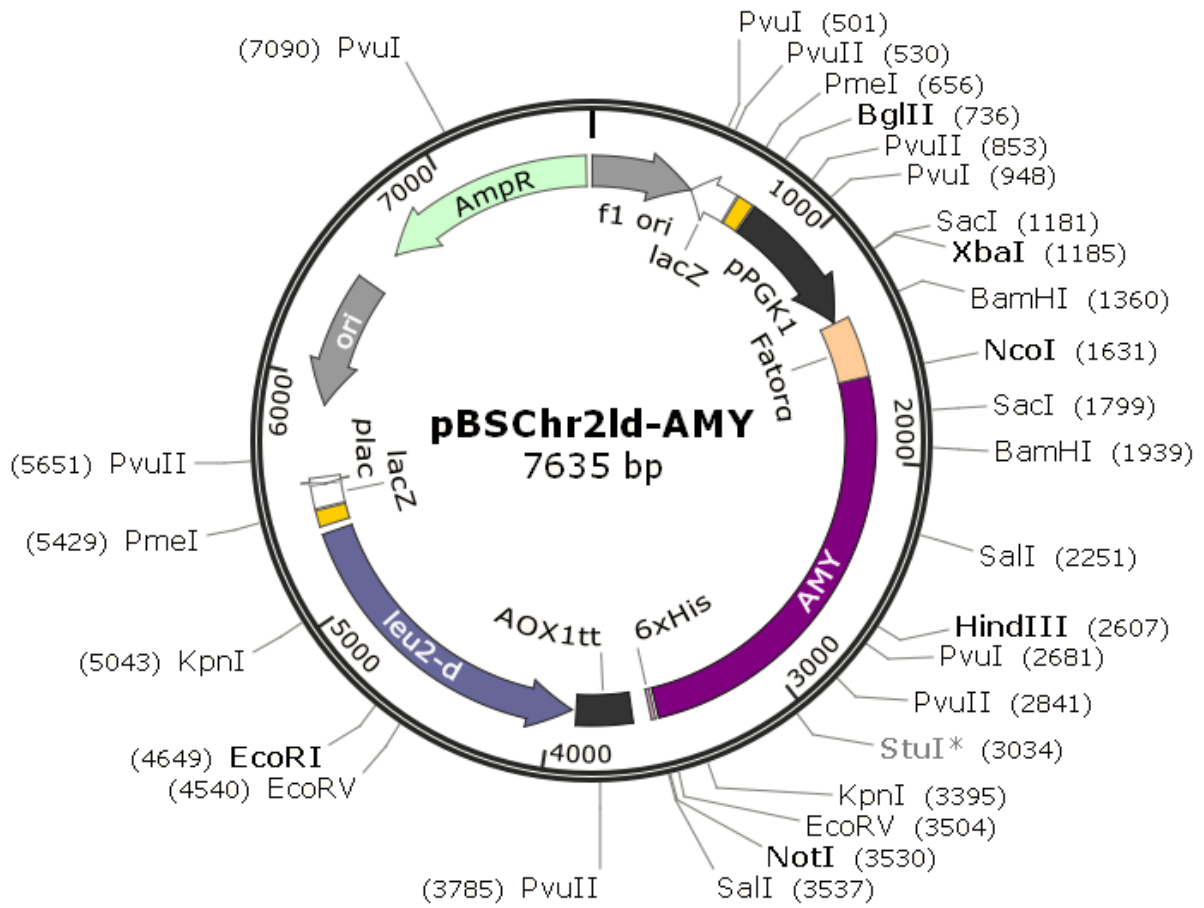
## Anexo 7: Mapa físico do vetor pBS5Sld-GFP



## Anexo 8: Mapa físico do vetor pPCVNTSId-GFP

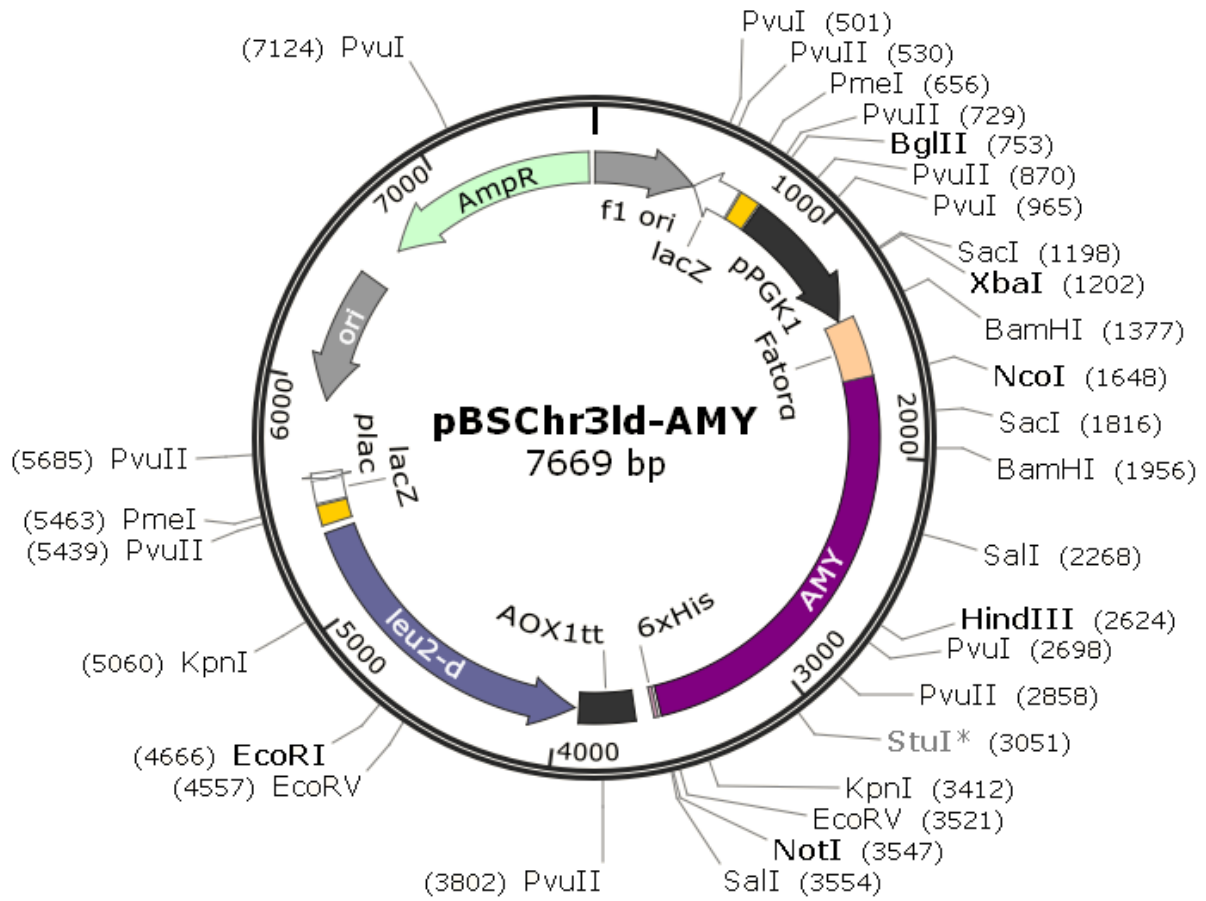


## Anexo 9: Mapa físico do vetor pBSChr2ld-AMY

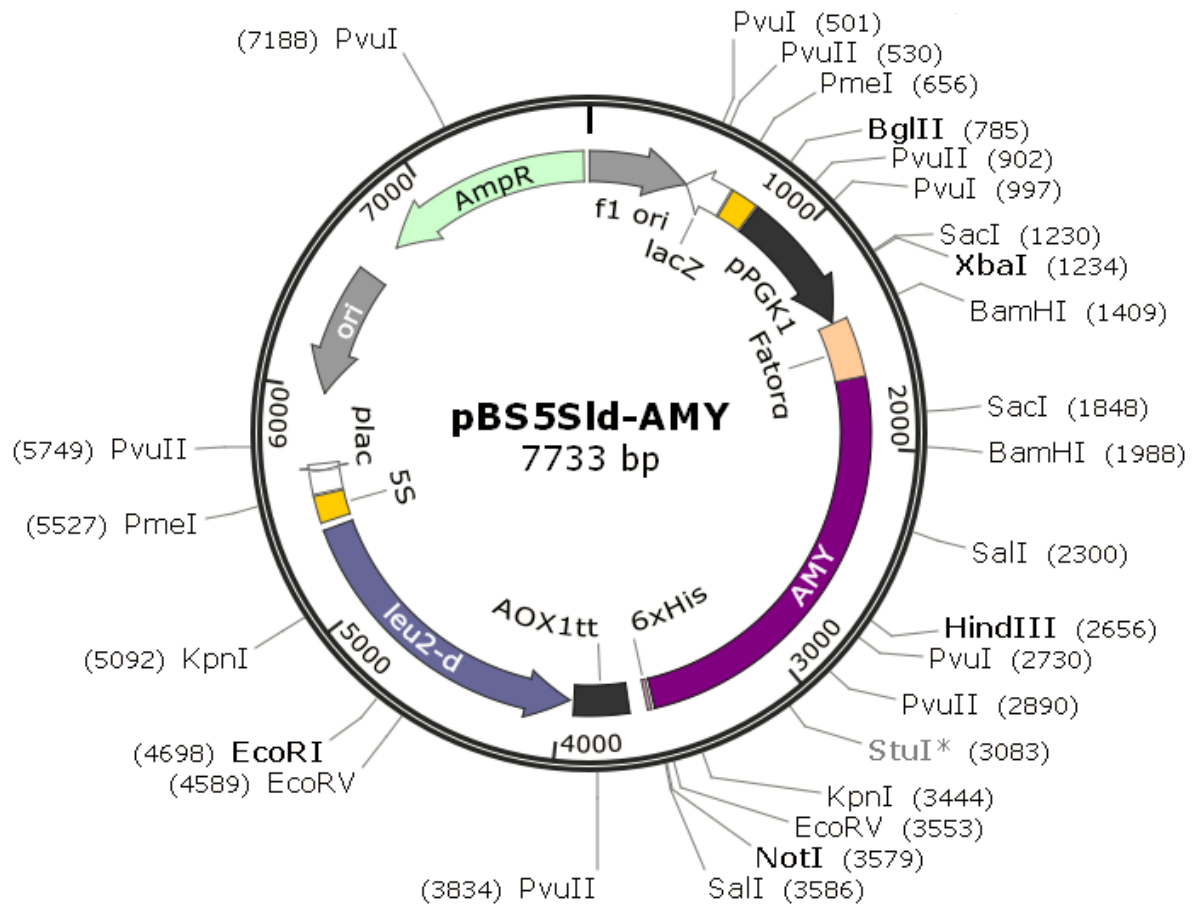




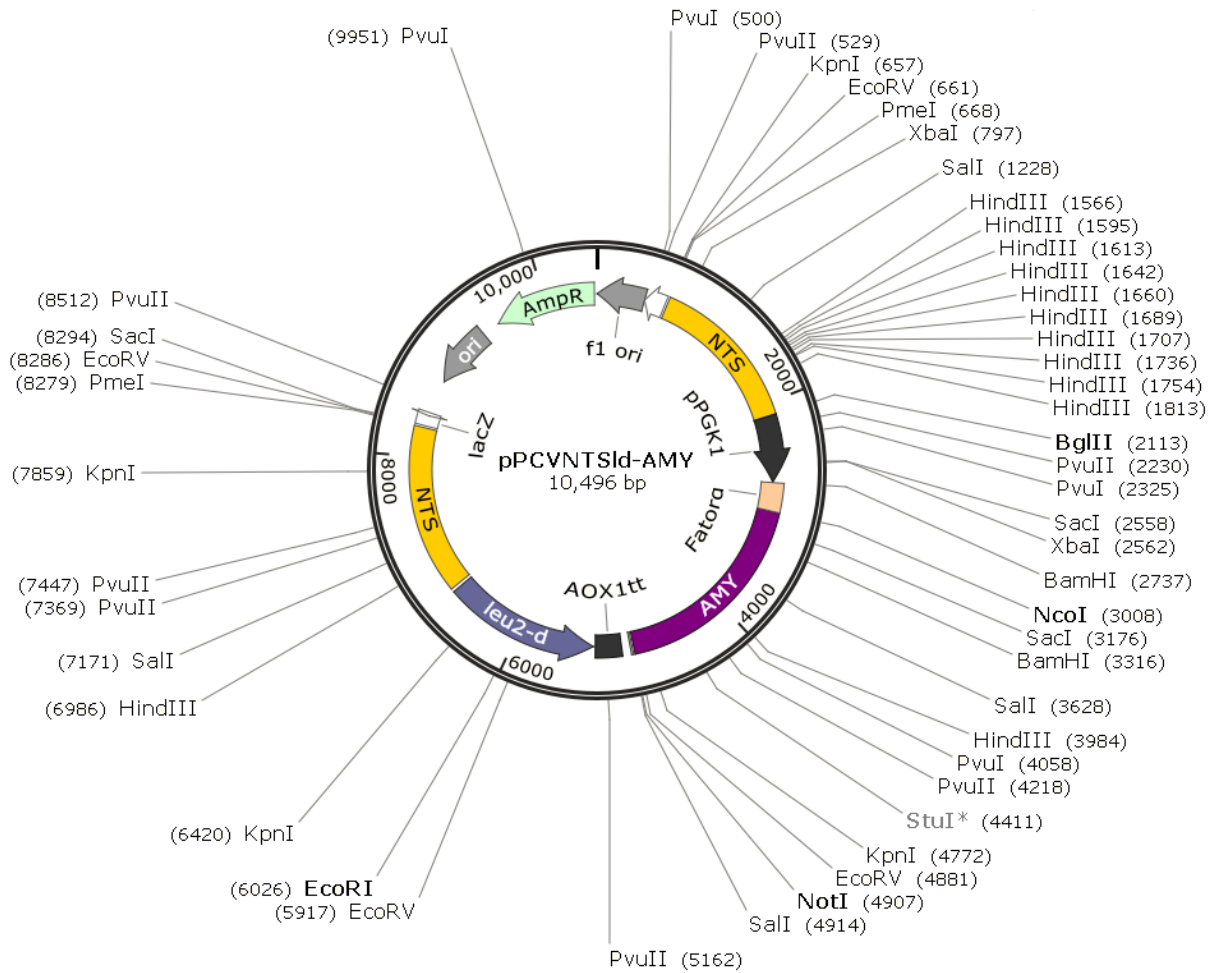
**Anexo 10: Mapa físico do vetor pBSChr3ld-AMY**



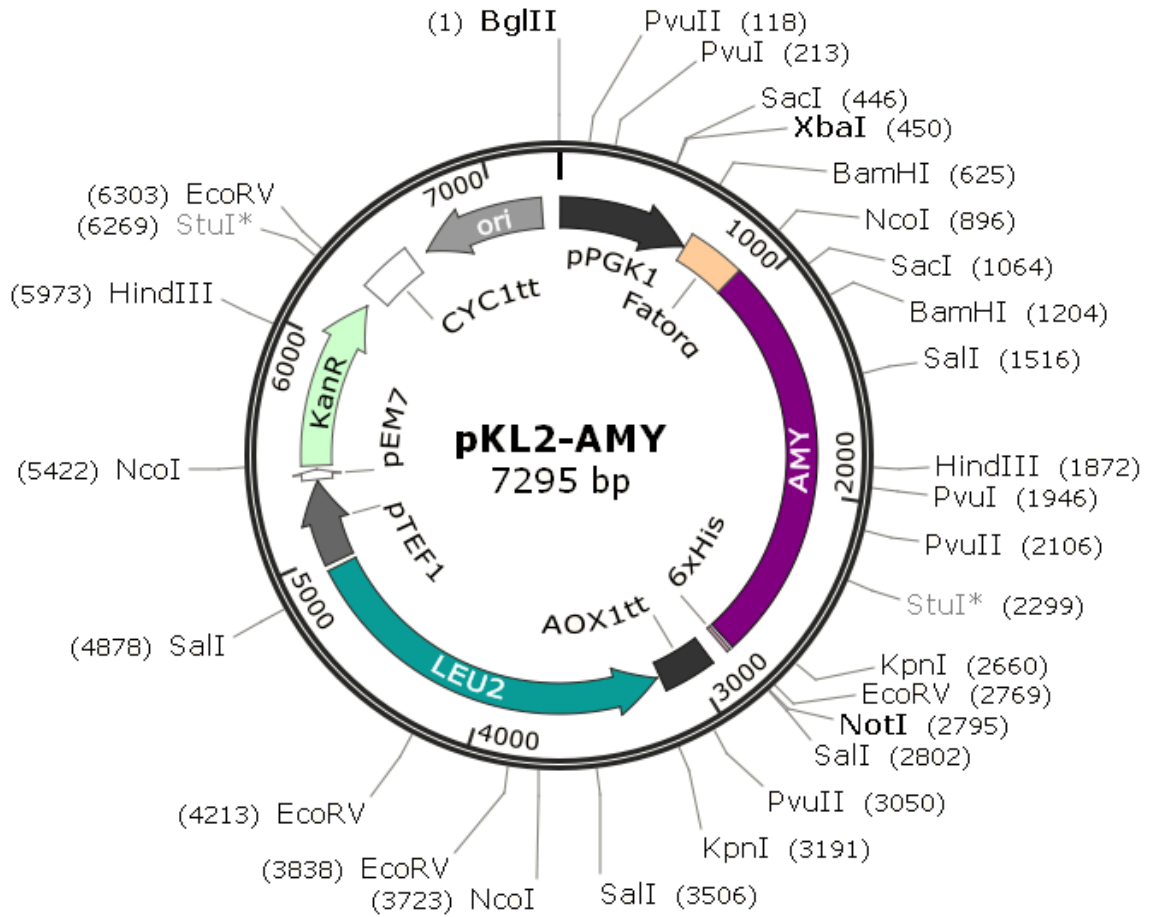
### Anexo 11: Mapa físico do vetor pBS5Sld-AMY



## Anexo 12: Mapa físico do vetor pPCVNTSId-AMY



### Anexo 13: Mapa físico do vetor pKL2-AMY



## 9. PATENTE E ARTIGOS PUBLICADOS



09/02/2017 870170008698  
10:19  
00.000.2.2.17.0065525.0

### Pedido nacional de Invenção, Modelo de Utilidade, Certificado de Adição de Invenção e entrada na fase nacional do PCT

Número do Processo: BR 10 2017 002626 4

#### Dados do Depositante (71)

---

Depositante 1 de 1

Nome ou Razão Social: FUNDAÇÃO UNIVERSIDADE DE BRASÍLIA/CENTRO DE APOIO AO DESENVOLVIMENTO TECNOLÓGICO

Tipo de Pessoa: Pessoa Jurídica

CPF/CNPJ: 00038174000143

Nacionalidade: Brasileira

Endereço: CENTRO DE APOIO AO DESENVOLVIMENTO TECNOLÓGICO - UNIVERSIDADE DE BRASÍLIA, CAMPUS UNIVERSITÁRIO DARCY RIBEIRO, EDIFÍCIO CDT, ASA NORTE

Cidade: Brasília

Estado: DF

CEP: 70904-970

País: Brasil

Telefone: (61)3107-4100/4128

Fax: (61)3107-4136

Email: nupitec@cdt.unb.br

## Dados do Pedido

---

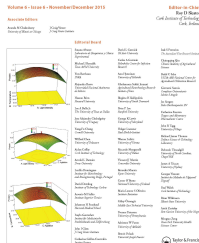
Natureza Patente: 10 - Patente de Invenção (PI)

Título da Invenção ou Modelo de SISTEMA DE EXPRESSÃO HETERÓLOGA EM LEVEDURA

Utilidade (54): AUXOTRÓFICA PARA LEUCINA CONTENDO VETOR DE EXPRESSÃO COM MARCA DE SELEÇÃO DEFECTIVA leu2-d, PROCESSO DE OBTENÇÃO DESSE SISTEMA E USO DE SEQUÊNCIAS REPETIDAS PARA INTEGRAÇÃO MÚLTIPLA

Resumo: A presente invenção refere-se à construção de vetor de expressão contendo a marca defectiva leu2-d, que permite a seleção de células transformantes contendo um maior número de cópias do vetor integradas no genoma. O escopo da invenção compreende um sistema de expressão contendo conjunto de vetores e linhagem de levedura auxotrófica, e o uso de sequências repetidas do cromossomo II, cromossomo III e rDNA 5S da levedura para promover recombinação homóloga nesses loci no genoma da levedura. A presença dessas sequências repetidas no vetor leva à sua integração múltipla. A linhagem auxotrófica para o aminoácido leucina é obtida após modificações genéticas em células de levedura para interromper o gene LEU2. As células modificadas são utilizadas como hospedeiras para a expressão de genes presentes em vetores contendo a marca de seleção leu2-d. A transformação da linhagem auxotrófica leu2 com os vetores da invenção possibilita a recuperação da prototrofia quando um alto número de cópias do vetor é integrado, expressando, dessa forma, a proteína codificada pelo gene de interesse inserido no vetor.

Figura a publicar: 10



# Molecular strategies to increase the levels of heterologous transcripts in *Komagataella phaffii* for protein production

Luiza Cesca Piva, Maritza Ocampo Bentacur, Viviane Castelo Branco Reis, Janice Lisboa De Marco, Lidia Maria Pepe de Moraes & Fernando Araripe Gonçalves Torres

To cite this article: Luiza Cesca Piva, Maritza Ocampo Bentacur, Viviane Castelo Branco Reis, Janice Lisboa De Marco, Lidia Maria Pepe de Moraes & Fernando Araripe Gonçalves Torres (2017): Molecular strategies to increase the levels of heterologous transcripts in *Komagataella phaffii* for protein production, *Bioengineered*, DOI: [10.1080/21655979.2017.1296613](https://doi.org/10.1080/21655979.2017.1296613)

To link to this article: <http://dx.doi.org/10.1080/21655979.2017.1296613>



Accepted author version posted online: 22 Feb 2017.  
Published online: 22 Feb 2017.



Submit your article to this journal [↗](#)



Article views: 4



View related articles [↗](#)



View Crossmark data [↗](#)

COMMENTARY

## Molecular strategies to increase the levels of heterologous transcripts in *Komagataella phaffii* for protein production

Luiza Cesca Piva<sup>†</sup>, Maritza Ocampo Bentacur<sup>†</sup>, Viviane Castelo Branco Reis, Janice Lisboa De Marco, Lidia Maria Pepe de Moraes, and Fernando Araripe Gonçalves Torres

Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF, Brazil

### ABSTRACT

*Komagataella phaffii* (formerly *Pichia pastoris*) is a well-known fungal system for heterologous protein production in the context of modern biotechnology. To obtain higher protein titers in this system many researchers have sought to optimize gene expression by increasing the levels of transcription of the heterologous gene. This has been typically achieved by manipulating promoter sequences or by generating clones bearing multiple copies of the desired gene. The aim of this work is to describe how these different molecular strategies have been applied in *K. phaffii* presenting their advantages and drawbacks.

### ARTICLE HISTORY

Received 10 February 2017  
Accepted 10 February 2017

### KEYWORDS

gene transcription;  
heterologous expression;  
*Komagataella phaffii*; multi-  
copy integration; promoter  
engineering

### Introduction

The methylotrophic yeast *Komagataella phaffii*, still mostly known by its old name, *Pichia pastoris*, is considered one of the most important platforms in modern biotechnology for the production of heterologous proteins. This is due to several features including its high volumetric productivity and ability to perform some posttranslational modifications in a manner similar to mammalian cells (for review see <sup>1,2</sup>). However, a clear limitation of this system is its low specific productivity which has prompted many researchers to pursue molecular strategies to improve protein production. The most popular approach to reach this goal is to offer to the translational machinery of the cell higher titers of a particular transcript. This is usually accomplished by driving the expression of the heterologous gene with a suitable promoter or by simply increasing the copy number of the target gene.<sup>3</sup> One should consider that gene overexpression may pose a metabolic burden to the physiology of the cell thus resulting in disappointing outputs. Nonetheless, most researchers still consider these approaches (summarized in Fig. 1) as a starting point in the endeavor of obtaining higher protein titers in *K. phaffii*.

### Inducible and constitutive promoters

*K. phaffii* vectors are derived from a few integrative plasmids which can carry expression cassettes under the control of inducible or constitutive promoters. A detailed review on the promoters available for protein production in *K. phaffii* is provided elsewhere.<sup>4</sup> The first expression system for *K. phaffii* was based on the promoter of the alcohol oxidase gene ( $P_{AOX1}$ ) involved in the first step of methanol metabolism. This strong and methanol-induced promoter was partly responsible for the popularity of this yeast as a heterologous protein production platform since the 1980s.<sup>5,6</sup>  $P_{AOX1}$  is repressed by glucose, glycerol and ethanol, therefore it represents a reliable tool when expressing toxic or growth-impairing proteins due to its tight regulation.<sup>7</sup> Several other promoters involved in the methanol utilization pathway have been isolated and classified as strong, intermediate or weak according to its expression levels in comparison to  $P_{AOX1}$ .<sup>8</sup> Because induction by methanol represents a hazardous disadvantage for large-scale processes alternative inducible promoters have been considered. A novel regulated promoter has recently been identified from a high-affinity glucose transporter gene (*GTH1*) which is repressed by

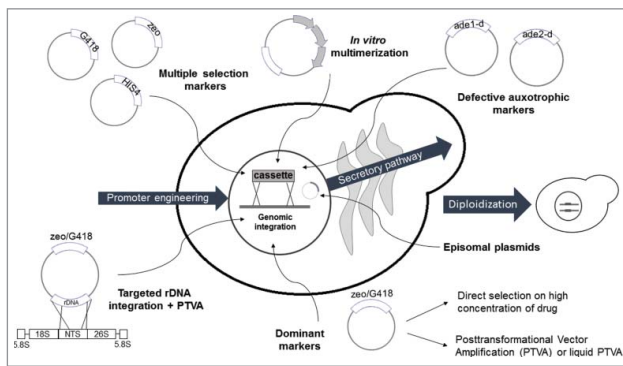
**CONTACT** Fernando Araripe Gonçalves Torres ✉ [ftorres@unb.br](mailto:ftorres@unb.br) 📧 Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF Brazil 70910-900.

Color versions of one or more of the figures in this article can be found online at [www.tandfonline.com/kbie](http://www.tandfonline.com/kbie)

<sup>†</sup>These authors contributed equally to this work.

© 2017 Taylor & Francis





**Figure 1.** Molecular strategies used in *K. phaffii* to improve gene transcription. The titer of heterologous transcripts can be improved by using strong promoters (engineered or not) or by generating clones with extra doses of the desired gene integrated in the genome or present in episomal plasmids. The extra-copies may be introduced by different approaches: sequential transformation with vectors bearing different markers, *in vitro* multimerization, colony screening under drug selective pressure, gene amplification by PTVA (or liquid PTVA), use of defective auxotrophic markers in minimal medium and diploidization of selected clones. Abbreviations: zeo/G418 = dominant markers which confer resistance to zeocin or G418, respectively.

glycerol.<sup>9</sup> Since  $P_{GTH1}$  is responsive to carbon source depletion it should represent an attractive and cheap alternative to other regulated systems that rely on the addition of an external source of the inducing agent. Recently, 2 inducible promoters from the *K. phaffii* rhamnose utilization pathway were proposed for the production of food-grade and therapeutically important recombinant proteins.<sup>10</sup> Also, extensive transcriptome analysis has led to the identification of new promoters regulated by the carbon source which should represent a fine addition to the *K. phaffii* promoter toolbox.<sup>11</sup>

Another alternative to methanol-induction is the use of constitutive expression systems which may be more advantageous when considering the fermentation strategy used. Whereas inducible expression requires a growth phase before the addition of methanol (production phase) in a constitutive system protein production is concomitant to cell growth, thus favoring continuous fermentation processes.<sup>12</sup> The promoter derived from the glycolytic glyceraldehyde 3-phosphate dehydrogenase gene ( $P_{GAP}$ ) is one of the most popular constitutive systems. Full transcription from  $P_{GAP}$  is achieved when grown in glucose as a carbon source.<sup>13,14</sup> In addition to  $P_{GAP}$  other moderately strong constitutive systems may be used such as  $P_{TEF1}$ <sup>15</sup> and  $P_{PGKI}$ .<sup>16,17</sup>

## Promoter engineering

A wide range of available promoters is an advantage for *K. phaffii* as a heterologous protein production platform since it allows the construction of fine-tuned systems and strains while also meeting requirements set by specific proteins.<sup>4</sup> For protein production purposes, intermediate or weak promoters may be more advantageous when protein folding becomes an issue as a result of gene overexpression. Likewise, in metabolic engineering applications a diverse set of promoters with different strengths allows a more accurate tuning of the expression levels and metabolic flux.<sup>3</sup> Transcription-factor binding sites (TFBS), upstream repression sites (URS) and upstream activation sequences (UAS) from different *K. phaffii* promoters may direct the design of new synthetic promoter sequences in the future.<sup>18</sup> Promoter libraries have already been constructed using not only random mutagenesis methods but also with specific synthetic core promoter sequences derived from wild-type promoter alignments.<sup>9,19</sup> The  $P_{GAP}$  sequence has also been manipulated through error-prone PCR for the construction of a  $GAP$ -promoter library with strengths varying from ~0,6% to 19,6-fold of that of the wild-type  $P_{GAP}$ .<sup>20</sup>

## Multi-copy number clones

One of the most common strategies to increase the levels of mRNA is by introducing several copies of the heterologous gene.<sup>3,21</sup> In several cases this has led to a significant improvement in heterologous protein production, most notably when considering intracellular expression.<sup>22</sup> However, in some cases involving secreted proteins, a higher gene copy number resulted in reduced heterologous protein production. For example, production of porcine insulin precursor (PIP) under inductive conditions was significantly affected in clones bearing >12 copies of the heterologous gene.<sup>23</sup> This has been attributed to bottlenecks in the secretory pathway which may become overwhelmed by the effects of high titers of a particular heterologous protein.<sup>24</sup> Strategies to overcome these barriers may include the use of new signal peptides based on the *K. phaffii* secretome<sup>25</sup> and combinatorial engineering of secretion helper factors involved in protein folding and vesicle trafficking.<sup>26</sup>

In *K. phaffii*, expression vectors are typically integrated into the host chromosome by homologous

recombination and non-homologous end joining events.<sup>27</sup> Multiple integration events may occur at low frequency (5–6%) by the integration of expression cassettes *in tandem* in a head-to-tail configuration.<sup>28</sup> These rare integration events can be favored by using vectors with dominant markers based on antibiotic resistance genes which allow the screening of desired clones under drug selective pressure. The use of dominant markers is the basis of the “posttransformational vector amplification” method (PTVA)<sup>29</sup> or its recently described variant “liquid PTVA”.<sup>30</sup> In PTVA, cells transformed with one or few copies of a vector carrying the marker that confers resistance to zeocin or G418 can be selected in higher concentrations of the antibiotic in a stepwise manner resulting in isolation of multi-copy clones several days after the initial transformation.

Multi-copy integration events can be enhanced by targeting vector insertion to repetitive sequences in the genome such as the rDNA cluster (rDNA) which comprises a 7450-bp repeated sequence organized *in tandem*. When a specific sequence is targeted to this cluster, a low copy number of integrated vectors is initially obtained, then, a strong selective pressure is applied for vector amplification.<sup>28</sup> The use of the non-transcribed spacer region (NTS) of the rDNA as an integration target in combination with the PTVA method has led to the successful isolation of multi-copy clones in *K. phaffii*.<sup>22</sup>

In addition to the high costs of eukaryotic antibiotics, an important drawback in the use of dominant markers is that in several clones an increase in drug resistance does not necessarily reflect multi-copy integration or an improvement in protein production.<sup>3,22</sup> Moreover, there are concerns that the accidental release of genetically modified organisms bearing drug-resistance markers may be horizontally transferred to environmental organisms.<sup>31</sup> Marker-removal has been successfully accomplished in *K. phaffii* with the use of the Cre/loxP<sup>32</sup> and FLP/FRT<sup>27</sup> site-specific recombination systems. The use of the *Escherichia coli* counter-selectable toxin gene *mazF* may further improve the screening of marker-free clones.<sup>33</sup> However, one should be aware that undesirable chromosomal rearrangement events may occur when using recombinase-based technology, especially when considering the simultaneous removal of multi-copy drug markers spread in the genome. The CRISPR-Cas9 system has recently been established in *K. phaffii*<sup>34</sup> and

may represent a powerful tool for marker-removal and generation of auxotrophic strains in this yeast.

The use of defective auxotrophic markers is an alternative to select multiple integration events. These markers generally represent biosynthetic pathway genes with a truncated promoter and, therefore, are transcribed at low levels. To compensate its low expression levels, transformed clones are selected to carry a high copy number of the defective marker to restore prototrophy. Defective markers have been extensively used in *S. cerevisiae* as a strategy to amplify expression plasmids.<sup>35,36</sup> In *K. phaffii*, the use of defective *ADE1* and *ADE2* alleles as selection markers has favored multiple integration with the concomitant increase in heterologous protein production.<sup>37</sup>

Although the use of defective markers for multi-copy integration in *K. phaffii* requires the development of new auxotrophic mutant strains and the use of specific media it presents some clear advantages. Most importantly, clones with different copies of the desired gene are easily obtained on a single transformation event without the need of laborious replica plating steps as required in PTVA. Also, it is a cheap alternative to drug selection and recombinant clones that carry auxotrophic marker do not pose any significant threat to the environment.

Other less frequently used methods to obtain multi-copy clones have been described. One involves the *in vitro* construction of multimers of the expression cassette.<sup>38</sup> In this case the size of the expression cassette may hamper DNA manipulation. Another method involves the sequential transformation of the host strain with different vectors bearing the expression cassettes, a laborious procedure which also requires the availability of strains with several genetic markers. However, the resulting strains may be further crossed to generate diploid cells thus increasing overall copy number of the desired sequence.<sup>39</sup>

Recently, an episomal plasmid carrying the panARS sequence was tested for recombinant protein production in *K. phaffii*.<sup>40</sup> This autonomously replicating sequence derived from *Kluyveromyces lactis* conferred stable replicative maintenance to plasmids and allowed the selection of clones with 6 to 19 copies of the plasmid. Since the use of episomal vectors represent an approach that leads to increased gene copy number with a higher efficiency of transformation and clonal homogeneity it should draw more interest in

the near future as a platform for heterologous protein production in *K. phaffii*.

## Conclusion

It is clear that the improvement on protein production as a result of the use of any of the molecular strategies described in this work should be assessed on a case-by-case basis. In our experience, the use of a vector bearing a repetitive target sequence combined with an auxotrophic defective marker represents an interesting start point for the development of an expression platform in *K. phaffii*.

## Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

## Funding

This work was supported by grants from Fundação de Amparo à Pesquisa do Distrito Federal (FAP/DF) and CNPq, Brazil.

## References

- [1] Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast* 2005; 22(4):249-270; PMID:15704221; <http://dx.doi.org/10.1002/yea.1208>
- [2] Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M, Mattanovich D. *Pichia pastoris*: protein production host and model organism for biomedical research. *Future Microbiol* 2013; 8(2):191-208; PMID:23374125; <http://dx.doi.org/10.2217/fmb.12.133>
- [3] Aw R, Polizzi KM. Can too many copies spoil the broth? *Microb Cell Fact* 2013; 12:128; PMID:24354594; <http://dx.doi.org/10.1186/1475-2859-12-128>
- [4] Vogl T, Glieder A. Regulation of *Pichia pastoris* promoters and its consequences for protein production. *Nature Biotechnol* 2013; 30(4):385-404.
- [5] Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Mol Cell Biol* 1985; 5(5):1111-1121; PMID:3889590; <http://dx.doi.org/10.1128/MCB.5.5.1111>
- [6] Cereghino GP, Cereghino JL, Ilgen C, Cregg JM. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr Opin Biotechnol* 2002; 13(4):329-332; PMID:12323354; [http://dx.doi.org/10.1016/S0958-1669\(02\)00330-0](http://dx.doi.org/10.1016/S0958-1669(02)00330-0)
- [7] Cereghino JL, Cregg JM. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 2000; 24(1):45-66; PMID:10640598; <http://dx.doi.org/10.1111/j.1574-6976.2000.tb00532.x>
- [8] Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl AM, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, et al. A toolbox of diverse promoters related to methanol utilization: functionally verified parts for heterologous pathway expression in *Pichia pastoris*. *ACS Synth Biol*. 2015; 5(2):172-186; PMID:26592304; <http://dx.doi.org/10.1021/acssynbio.5b00199>
- [9] Prielhofer R, Maurer M, Klein J, Wenger J, Kiziak C, Gasser B, Mattanovich D. Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. *Microb Cell Fact* 2013; 12:5; PMID:23347568; <http://dx.doi.org/10.1186/1475-2859-12-5>
- [10] Liu B, Zhang Y, Zhang X, Yan C, Zhang Y, Xu X, Zhang W. Discovery of a rhamnose utilization pathway and rhamnose-inducible promoters in *Pichia pastoris*. *Sci Rep* 2016; 6:27352; PMID:27256707; <http://dx.doi.org/10.1038/srep27352>
- [11] Love KR, Shah KA, Whittaker CA, Wu J, Bartlett MC, Ma D, Leeson RL, Priest M, Borowsky J, Young SK, Love JC. Comparative genomics and transcriptomics of *Pichia pastoris*. *BMC Genom* 2016; 17:550; <http://dx.doi.org/10.1186/s12864-016-2876-y>
- [12] Vassileva A, Chugh DA, Swaminathan S, Khanna N. Expression of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris* using the GAP promoter. *J Biotechnol* 2001; 88(1):21-35; PMID:11377762; [http://dx.doi.org/10.1016/S0168-1656\(01\)00254-1](http://dx.doi.org/10.1016/S0168-1656(01)00254-1)
- [13] Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM. Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 1997; 186(1):37-44; PMID:9047342; [http://dx.doi.org/10.1016/S0378-1119\(96\)00675-0](http://dx.doi.org/10.1016/S0378-1119(96)00675-0)
- [14] Zhang AL, Luo JX, Zhang TY, Pan YW, Tan YH, Fu CY, Tu FZ. Recent advances on the GAP promoter derived expression system of *Pichia pastoris*. *Mol Biol Rep* 2009; 36(6):1611-1619; PMID:18781398; <http://dx.doi.org/10.1007/s11033-008-9359-4>
- [15] Ahn J, Hong J, Lee H, Park M, Lee E, Kim C, Choi E, Jung J. Translation elongation factor 1-alpha gene from *Pichia pastoris*: molecular cloning, sequence, and use of its promoter. *Appl Microbiol Biotechnol* 2007; 74(3):601-608; PMID:17124582; <http://dx.doi.org/10.1007/s00253-006-0698-6>
- [16] de Almeida JR, de Moraes LM, Torres FA. Molecular characterization of the 3-phosphoglycerate kinase gene (*PGK1*) from the methylotrophic yeast *Pichia pastoris*. *Yeast* 2005; 22(9):725-737; PMID:16034819; <http://dx.doi.org/10.1002/yea.1243>
- [17] Arruda A, Reis VC, Batista VD, Daher BS, Piva LC, De Marco JL, de Moraes LM, Torres FA. A constitutive expression system for *Pichia pastoris* based on the *PGK1* promoter. *Biotechnol Lett* 2016; 38(3):509-517; PMID:26585331; <http://dx.doi.org/10.1007/s10529-015-2002-2>
- [18] Hartner FS, Ruth C, Langenegger D, Johnson SN, Hyka P, Lin-Cereghino GP, Lin-Cereghino J, Kovar K, Cregg

- JM, Glieder A. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res* 2008; 36(12):e76; PMID:18539608; <http://dx.doi.org/10.1093/nar/gkn369>
- [19] Vogl T, Ruth C, Pitzer J, Kickenweiz T, Glieder A. Synthetic core promoters for *Pichia pastoris*. *ACS Synth Biol* 2014; 3(3):188-191; PMID:24187969; <http://dx.doi.org/10.1021/sb400091p>
- [20] Qin X, Qian J, Yao G, Zhuang Y, Zhang S, Chu J. GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. *Appl Environ Microbiol* 2011; 77(11):3600-3608; PMID:21498769; <http://dx.doi.org/10.1128/AEM.02843-10>
- [21] Romanos M, Scorer C, Sreerikshna K, Clare J. The generation of multicopy recombinant strains. *Methods Mol Biol* 1998; 103:55-72; PMID:9680633
- [22] Marx H, Mecklenbrauker A, Gasser B, Sauer M, Mattanovich D. Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. *FEMS Yeast Res* 2009; 9(8):1260-1270; PMID:19799640; <http://dx.doi.org/10.1111/j.1567-1364.2009.00561.x>
- [23] Zhu T, Guo M, Tang Z, Zhang M, Zhuang Y, Chu J, Zhang S. Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast *Pichia pastoris*. *J Appl Microbiol* 2009; 107(3):954-963; PMID:19486418; <http://dx.doi.org/10.1111/j.1365-2672.2009.04279.x>
- [24] Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnol Bioeng* 2004; 85(4):367-375; PMID:14755554; <http://dx.doi.org/10.1002/bit.10904>
- [25] Massahi A, Calik P. In-silico determination of *Pichia pastoris* signal peptides for extracellular recombinant protein production. *J Theor Biol* 2015; 364:179-188; PMID:25218497; <http://dx.doi.org/10.1016/j.jtbi.2014.08.048>
- [26] Guan B, Chen F, Su S, Duan Z, Chen Y, Li H, Jin J. Effects of co-overexpression of secretion helper factors on the secretion of a HSA fusion protein (IL2-HSA) in *Pichia pastoris*. *Yeast* 2016; 33(11):587-600; <http://dx.doi.org/10.1002/yea.3183>
- [27] Naatsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One* 2012; 7(6):e39720; PMID:22768112; <http://dx.doi.org/10.1371/journal.pone.0039720>
- [28] Lopes TS, Hakkaart GJ, Koerts BL, Raue HA, Planta RJ. Mechanism of high-copy-number integration of pMIRY-type vectors into the ribosomal DNA of *Saccharomyces cerevisiae*. *Gene* 1991; 105(1):83-90; PMID:1937009; [http://dx.doi.org/10.1016/0378-1119\(91\)90516-E](http://dx.doi.org/10.1016/0378-1119(91)90516-E)
- [29] Sunga AJ, Tolstorukov I, Cregg JM. Posttransformational vector amplification in the yeast *Pichia pastoris*. *FEMS Yeast Res* 2008; 8(6):870-876; PMID:18637138; <http://dx.doi.org/10.1111/j.1567-1364.2008.00410.x>
- [30] Aw R, Polizzi KM. Liquid PTVA: a faster and cheaper alternative for generating multi-copy clones in *Pichia pastoris*. *Microb Cell Fact* 2016; 15:29; PMID:26849882; <http://dx.doi.org/10.1186/s12934-016-0432-8>
- [31] Droge M, Puhler A, Selbitschka W. Horizontal gene transfer as a biosafety issue: a natural phenomenon of public concern. *J Biotechnol* 1998; 64(1):75-90; PMID:9823660; [http://dx.doi.org/10.1016/S0168-1656\(98\)00105-9](http://dx.doi.org/10.1016/S0168-1656(98)00105-9)
- [32] Marx H, Mattanovich D, Sauer M. Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*. *Microb Cell Fact* 2008; 7:23; PMID:18664246; <http://dx.doi.org/10.1186/1475-2859-7-23>
- [33] Yang J, Jiang W, Yang S. mazF as a counter-selectable marker for unmarked genetic modification of *Pichia pastoris*. *FEMS Yeast Res* 2009; 9(4):600-9; PMID:19416369; <http://dx.doi.org/10.1111/j.1567-1364.2009.00503.x>
- [34] Weninger A, Hatzl AM, Schmid C, Vogl T, Glieder A. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J Biotechnol* 2016; 235:139-49; PMID:27015975; <http://dx.doi.org/10.1016/j.jbiotec.2016.03.027>
- [35] Kazemi Seresht A, Norgaard P, Palmqvist EA, Andersen AS, Olsson L. Modulating heterologous protein production in yeast: the applicability of truncated auxotrophic markers. *Appl Microbiol Biotechnol* 2013; 97(9):3939-3948; PMID:22782252; <http://dx.doi.org/10.1007/s00253-012-4263-1>
- [36] Moon HY, Lee DW, Sim GH, Kim HJ, Hwang JY, Kwon MG, Kang BK, Kim JM, Kang HA. A new set of rDNA-NTS-based multiple integrative cassettes for the development of antibiotic-marker-free recombinant yeasts. *J Biotechnol* 2016; 233:190-199; PMID:27411901; <http://dx.doi.org/10.1016/j.jbiotec.2016.07.006>
- [37] Du M, Battles MB, Nett JH. A color-based stable multi-copy integrant selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as auxotrophic markers. *Bioeng Bugs* 2012; 3(1):32-37; PMID:22126802
- [38] Li YT, Li MT, Fu CH, Zhou PP, Liu JM, Yu LJ. Improvement of arachidonic acid and eicosapentaenoic acid production by increasing the copy number of the genes encoding fatty acid desaturase and elongase into *Pichia pastoris*. *Biotechnol Lett* 2009; 31(7):1011-1017; PMID:19306085; <http://dx.doi.org/10.1007/s10529-009-9970-z>
- [39] Chen MT, Lin S, Shandil I, Andrews D, Stadheim TA, Choi BK. Generation of diploid *Pichia pastoris* strains by mating and their application for recombinant protein production. *Microb Cell Fact* 2012; 11:91; PMID:22748191; <http://dx.doi.org/10.1186/1475-2859-11-91>
- [40] Camattari A, Goh A, Yip LY, Tan AH, Ng SW, Tran A, Liu G, Liachko I, Dunham MJ, Rancati G. Characterization of a panARS-based episomal vector in the methylotrophic yeast *Pichia pastoris* for recombinant protein production and synthetic biology applications. *Microb Cell Fact* 2016; 15(1):139; PMID:27515025; <http://dx.doi.org/10.1186/s12934-016-0540-5>

TECHNICAL NOTES

Open Access



# Multicopy plasmid integration in *Komagataella phaffii* mediated by a defective auxotrophic marker

Maritza Ocampo Betancur<sup>1</sup>, Viviane Castelo Branco Reis<sup>1</sup>, André Moraes Nicola<sup>2</sup>, Janice Lisboa De Marco<sup>1</sup>, Lídia Maria Pepe de Moraes<sup>1</sup> and Fernando Araripe Gonçalves Torres<sup>1\*</sup>

## Abstract

**Background:** A commonly used approach to improve recombinant protein production is to increase the levels of expression by providing extra-copies of a heterologous gene. In *Komagataella phaffii* (*Pichia pastoris*) this is usually accomplished by transforming cells with an expression vector carrying a drug-resistance marker following a screening for multicopy clones on plates with increasingly higher concentrations of an antibiotic. Alternatively, defective auxotrophic markers can be used for the same purpose. These markers are generally transcriptionally impaired genes lacking most of the promoter region. Among the defective markers commonly used in *Saccharomyces cerevisiae* is *leu2-d*, an allele of *LEU2* which is involved in leucine metabolism. Cells transformed with this marker can recover prototrophy when they carry multiple copies of *leu2-d* in order to compensate the poor transcription from this defective allele.

**Results:** A *K. phaffii* strain auxotrophic for leucine (M12) was constructed by disrupting endogenous *LEU2*. The resulting strain was successfully transformed with a vector carrying *leu2-d* and an EGFP (enhanced green fluorescent protein) reporter gene. Vector copy numbers were determined from selected clones which grew to different colony sizes on transformation plates. A direct correlation was observed between colony size, number of integrated vectors and EGFP production. By using this approach we were able to isolate genetically stable clones bearing as many as 20 integrated copies of the vector and with no significant effects on cell growth.

**Conclusions:** In this work we have successfully developed a genetic system based on a defective auxotrophic which can be applied to improve heterologous protein production in *K. phaffii*. The system comprises a *K. phaffii leu2* strain and an expression vector carrying the defective *leu2-d* marker which allowed the isolation of multicopy clones after a single transformation step. Because a linear correlation was observed between copy number and heterologous protein production, this system may provide a simple approach to improve recombinant protein productivity in *K. phaffii*.

**Keywords:** *Komagataella phaffii*, Leucine biosynthesis, Auxotrophic marker, Multicopy integration, Expression system

## Background

The methylotrophic yeast *Komagataella phaffii* (formerly *Pichia pastoris*) is one of the most important expression platforms for the production of recombinant proteins [1, 2]. It offers many advantages such as: easy genetic manipulation; growth at high cell densities, e.g. 200 g L<sup>-1</sup> dry

weight during a glucose-limited fed-batch cultivation [3]; ability to produce heterologous proteins at high levels, e.g. more than 18 g L<sup>-1</sup> of lignocellulolytic enzyme TrCBH2 [4]; and post-translational modifications similar to higher eukaryotes [5].

Due to its biotechnological interest, many studies have focused on the genetic improvement of *K. phaffii* in order to optimize protein production. A well-established approach to accomplish this is to assure high transcription levels of a heterologous gene thus favoring the translation of the desired mRNA. Typically, this can be

\*Correspondence: ftorres@unb.br

<sup>1</sup> Laboratório de Biologia Molecular, Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF 70910-900, Brazil

Full list of author information is available at the end of the article

achieved by constructing expression cassettes under the control of strong promoters or/and by screening clones bearing multiple copies of the desired gene (for a review see [6, 7]). Genetic strategies are available for the isolation of multicopy clones. Yeast cells can be transformed with vectors carrying extra copies of the expression cassette cloned *in tandem* (multimeric construction) [8] or successive rounds of transformation can be performed using different selection markers [9]. In both cases cloning is labor-intensive and the extent of copy number increase is limited [10].

Another option consists in the use of antibiotic-resistance markers, in which case one looks for transformants growing in higher concentrations of the antibiotic (direct selection method) [11]. A previous study showed that this type of selection resulted in the isolation of sporadic multicopy integrants with increased productivity of the desired protein [12]. Dominant markers can also give rise to multicopy clones by posttransformational vector amplification (PTVA) [13] or liquid PTVA [14]. It has been demonstrated that after transformation with a few copies of a vector carrying a drug-resistance marker, such as zeocin or G418, cells can be selected in stepwise higher concentrations of the drug resulting in the selection of multicopy clones. The use of PTVA in combination with the use of rDNA non-transcribed sequence (NTS) as an integration target sequence resulted in multicopy clones in *K. phaffii* [15]. Besides being a laborious and expensive method due to the high costs of eukaryotic antibiotics, one disadvantage of the use of dominant markers is that a significant number of clones show increased natural drug-resistance for other reasons than vector copy number.

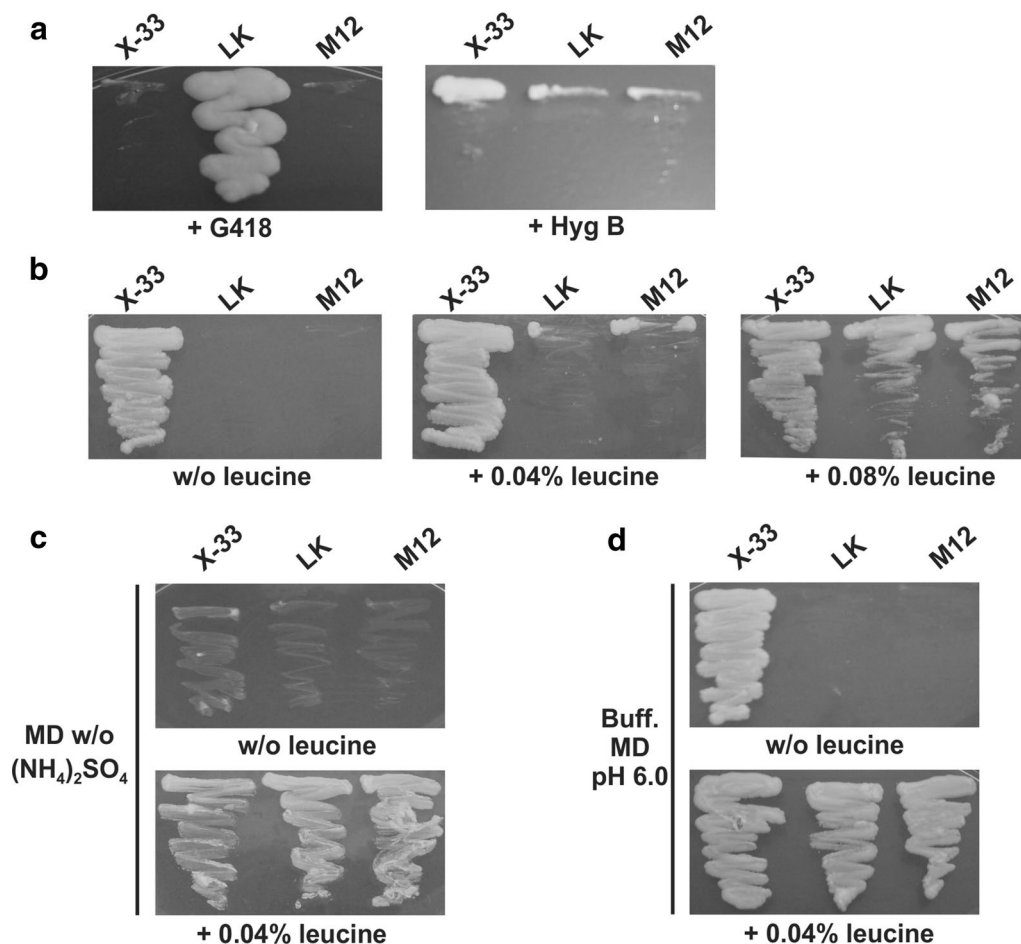
An alternative strategy is based in the use of defective auxotrophic markers, i.e. genes that are poorly transcribed typically due to extensive deletions of their promoters. To compensate the low transcription levels, cells need to amplify the copy number of the defective marker in order to recover prototrophy. Consequently, copy number of the neighboring heterologous gene is also amplified [16]. An example of such defective marker is the *leu2-d* allele which contains only 29 base pairs of the original promoter and is commonly used in *S. cerevisiae* for plasmid maintenance at high copy number under selective pressure [17]. Due to this feature, this system has also been used to increase recombinant protein production in this yeast [18–20]. This prompted us to develop an analogous system to be employed in *K. phaffii*. To accomplish this, we sought the construction of a *K. phaffii* strain auxotrophic for leucine and the development of an integrative expression vector based on *leu2-d* as a tool to increase recombinant protein production in this yeast.

## Results and discussion

### Construction of a *leu2* auxotrophic strain

Genetic manipulation of *K. phaffii* is possible due to its widely used transformation system which enables integration of foreign DNA into the genome via homologous recombination [21]. This approach has been successfully used to disrupt several genes in order to create auxotrophic mutants, e.g. *URA5* [22], *ARG1*, *ARG2*, *ARG3*, *HIS1*, *HIS2*, *HIS5* and *HIS6* [23]. Recently, a CRISPR-Cas9 system was developed for *K. phaffii* which has greatly facilitated gene knock out in this yeast [24].

We sought the development of a leucine auxotrophic strain by gene knock out of the endogenous *K. phaffii* *LEU2* gene with a *leu2::kan* disruption cassette. The resulting strain (LK) was then transformed with a plasmid expressing CreA recombinase for marker removal thus generating strain M12 (see Additional file 1 for details). Growth analysis on plates containing G418 or hygromycin B confirmed the loss of all dominant markers (Fig. 1a). The phenotypic behavior of the strains obtained with respect to leucine assimilation was then analyzed. As expected, LK and M12 strains were not able to grow on MD lacking leucine (Fig. 1b). We expected that supplementation of MD medium with leucine would allow both *leu2* strains to recover prototrophy, however, even with an oversupplementation (0.08%) of leucine cells were unable to grow (Fig. 1b). We reasoned that ammonium sulphate present in MD medium could be affecting leucine uptake because when this salt was replaced by 0.04% leucine as sole nitrogen source both *leu2* strains grew as well as wild-type X-33 (Fig. 1c). This result is in accordance with a previous work [25] which showed that cells grown in minimal medium exhibited an increase in the rate of leucine uptake when this amino acid was the sole nitrogen source. It is known that the addition of  $\text{NH}_4^+$  to yeast cells causes nitrogen catabolite inactivation and repression of several enzymes and permeases involved in the utilization of secondary nitrogen sources [26]. Leucine has been shown to be transported by at least three systems in *S. cerevisiae*: GAP (general amino acid permease), S1 (high-affinity permease) and S2 (low-affinity permease) [27]. In  $\text{NH}_4^+$ -containing media the activity of GAP is inhibited [28, 29] and the activity of S1 and S2 proteins is strongly reduced [30]. In addition, two redundant low affinity leucine permeases coded by the *AGP2* and *AGP3* genes are overexpressed when other permeases are inhibited [31]. The observation that prototrophy could only be achieved in high leucine concentrations can be explained by the fact that, in the presence of  $\text{NH}_4^+$ , leucine uptake is mainly due to low-affinity permeases. The effects of  $\text{NH}_4^+$  on leucine permeases could be related to intracellular pH as it has been shown that a *S. cerevisiae leu2* strain was more sensitive to internal



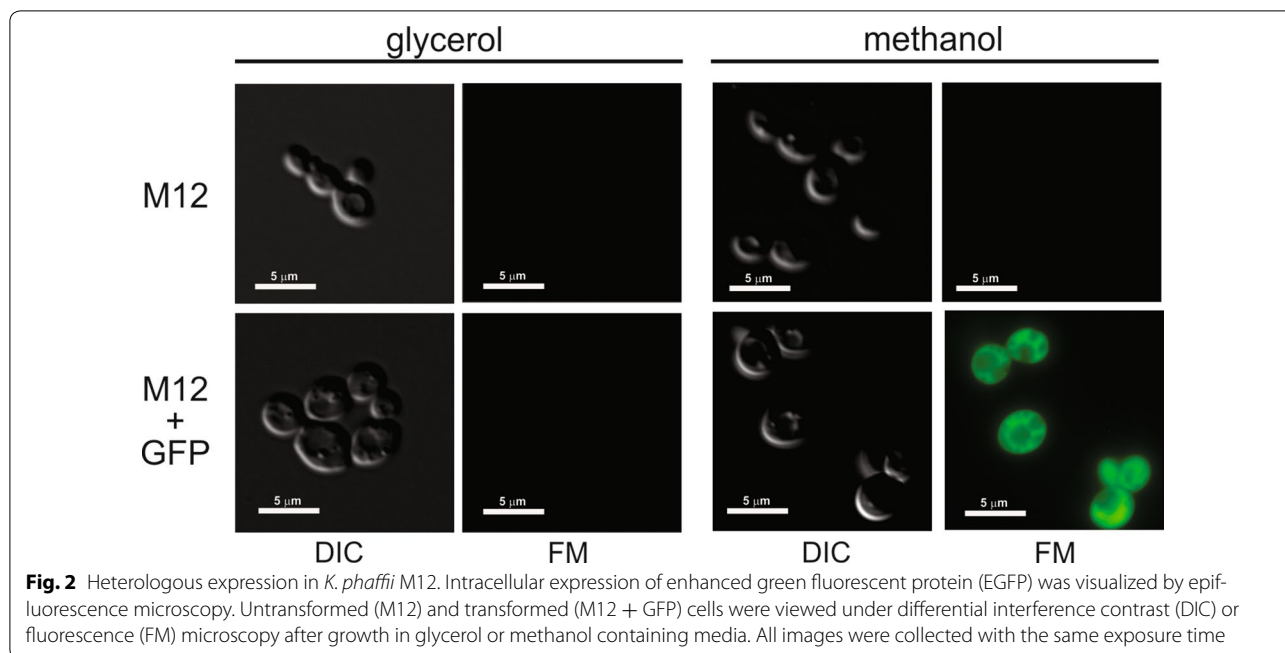
**Fig. 1** Strain phenotypic analysis. Strains LK and M12 were grown in different media to confirm drug resistance and leucine assimilation. X-33 was used as a control. **a** YPD containing G418 or hygromycin B (Hyg B). **b** MD lacking or not leucine in different concentrations. **c** MD with leucine replacing ammonium sulfate. **d** Buffered MD (pH 6.0) containing leucine. X-33 wild-type strain with intact *LEU2* gene, LK *leu2* strain disrupted with *kan* cassette, M12 strain obtained after marker removal with CreA recombinase

acidic conditions [32]. In yeast, amino acids and other nutrients are taken up by a proton symport mechanism [33]. When  $\text{NH}_4^+$  (a conjugated weak acid) is internalized by specific transporters it undergoes deprotonation and as a result the proton gradient is dissipated leading to acidification of the cytosol [34]. To test this, MD medium was buffered to pH 6.0 and as a result prototrophy was readily recovered as shown in Fig. 1d.

#### Heterologous expression in *K. phaffii* M12

In order to test *K. phaffii* M12 for heterologous expression we constructed an integrative vector, pGFP-L2, containing *LEU2* as selectable marker and the enhanced GFP (EGFP) reporter gene under the control of the methanol-inducible  $P_{AOX1}$  promoter. After electroporation of *K. phaffii* M12 with linearized pGFP-L2, one colony was randomly chosen and grown under non-inducible (glycerol) and inducible (methanol) conditions

to test for intracellular production of the EGFP. Figure 2 shows the result of fluorescence microscopy analysis for the presence of EGFP. As expected, no fluorescence was detected under non-inducible conditions and EGFP could only be detected when methanol was added. This result shows that *K. phaffii* M12 can recover prototrophy upon transformation with a vector carrying a wild-type *LEU2* marker. However, since the strategy used for *LEU2* disruption in M12 left the promoter region intact, a single copy of a defective version of this gene could potentially integrate at this *locus* by homologous recombination and reestablish prototrophy without the need of other copies of the marker. In order to develop a system based on defective *LEU2* for multicopy integration, we decided to use the *S. cerevisiae leu2-d* allele which has 68% identity with the *K. phaffii LEU2* homologue and should reduce the chances of homologous recombination at the disrupted *leu2 locus* in M12.

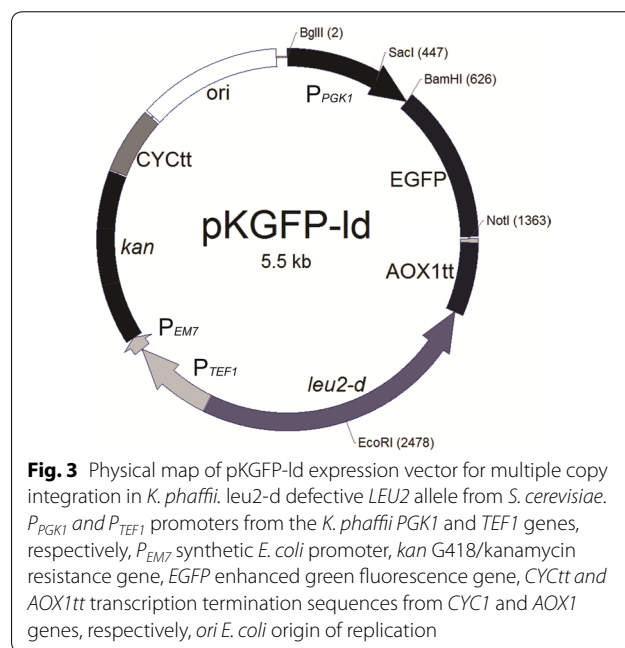


**Multiple copy integration**

Although the defective *leu2-d* allele has been successfully used to increase plasmid copy number and protein production in *S. cerevisiae* it has not yet been tested for the same purpose in *K. phaffii*. In a previous work, attenuated *ADE1* and *ADE2* genes involved in adenine biosynthesis were used to develop a color-based system for the screening of multicopy integrants in *K. phaffii* [35]. However, this system is based on large plasmids and, in some cases, the effects of background transcription from vector sequences were responsible for the recovery of adenine prototrophy.

We constructed an expression vector, pKGFP-Id (Fig. 3), carrying the *leu2-d* defective marker and an EGFP reporter construct under the control of the phosphoglycerate kinase 1 promoter ( $P_{PGK1}$ ). In order to circumvent the possibility of spurious marker expression, *leu2-d* was cloned in the opposite orientation of the other yeast promoters present on the vector ( $P_{PGK1}$  and  $P_{TEF1}$ ). Furthermore, the presence of the *kan* marker under the control of dual-promoters allows plasmid selection in *Escherichia coli* and may be used to confirm multicopy integration by plating transformed cells on media containing increasing concentrations of the antibiotic G418.

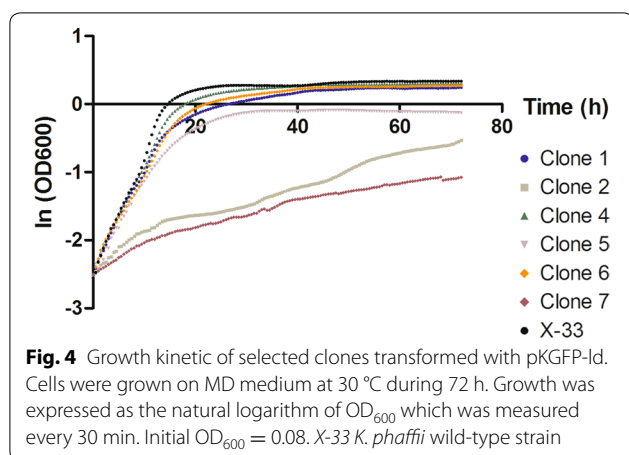
Seven days after transformation of *K. phaffii* M12 cells with pKGFP-Id, colonies of diverse sizes were observed on MD plates. Ten transformants representing colonies with different sizes were selected for further analysis. After a few passages on fresh MD plates four clones derived from the smallest colonies present on the original



transformation did not grown upon replica plating and were removed from this study. We speculate that these abortive clones were transformed with a limited number of copies of the defective marker, thus they were unable to sustain growth under selective conditions.

A growth kinetic analysis of the six remaining clones was performed on MD medium (Fig. 4). The calculated maximum growth rates (Table 1) showed that larger





**Table 1** Main features of the selected clones studied in this work

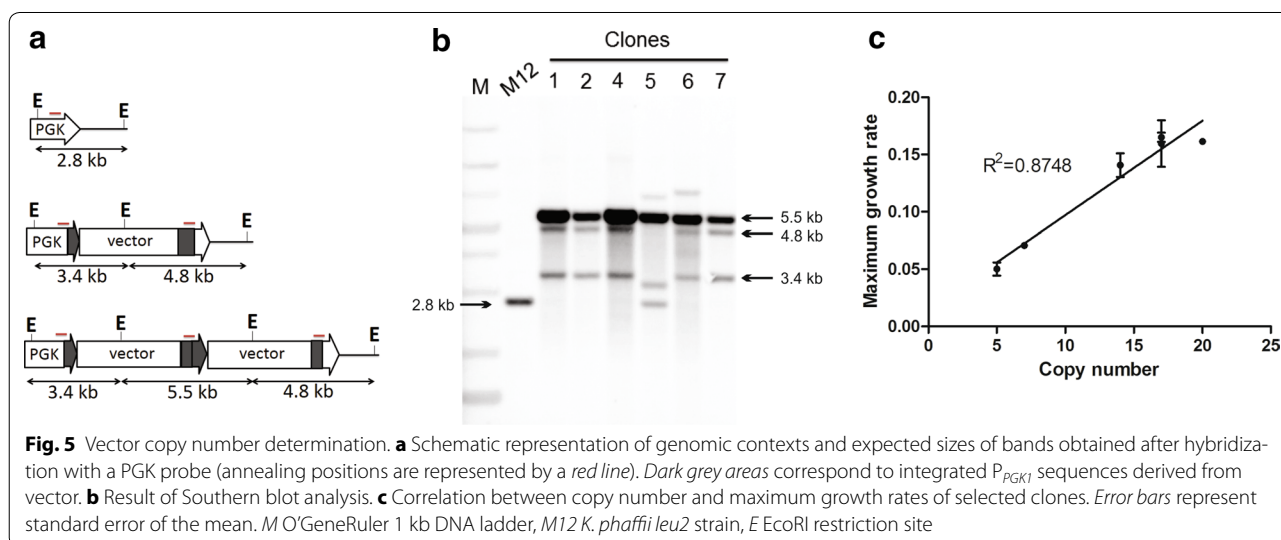
Clone	$\mu_{max}$ (h <sup>-1</sup> )	Vector copy number	EGFP production (fluorescence units)
1	0.1596 ± 0.0062	17	19215 ± 2388
2	0.0708 ± 0.0022	7	7705 ± 550
4	0.1614 ± 0.0027	20	23,688 ± 2152
5	0.1408 ± 0.0075	14	18,387 ± 1045
6	0.1650 ± 0.0078	17	20,074 ± 1738
7	0.0502 ± 0.0064	5	6711 ± 206
X-33	0.1707 ± 0.0069	0	ND

$\mu_{max}$  maximum growth rate, ND not determined

clones present on the original transformation plate (clones 1, 4, 5 and 6) exhibited a growth profile similar to that of *K. phaffii* X-33 whereas small sized colonies

(clones 2 and 7), as expected, exhibited smaller growth rates. Analysis of variance followed by Tukey’s post-test showed significant difference in maximum growth rates presented in Table 1 for clones 2 and 7 when compared to X-33 ( $p < 0.05$ ). Similar results were obtained when truncated *ADE1* and *ADE2* were used as selectable defective markers to transform *K. phaffii* [35]. In this case, multi-copy clones were also identified as colonies with larger sizes. We hypothesized that, in order to compensate the poor transcription from *leu2-d*, cells would require additional integrated copies of defective marker to recover full prototrophy. This prompted us to determine the copy number of vectors integrated into the *K. phaffii* genome by Southern blot.

Because pKGFp-I<sub>d</sub> was linearized with *Sac*I, a restriction site within  $P_{PGK1}$  (Fig. 3), most events of integration would be primarily targeted to this locus. According to the schematic representation shown in Fig. 5a, if a single copy of the vector had integrated into the *PGK1* locus, two fragments of 3.4 and 4.8 kb would be expected. Two or more integrated copies would yield an additional 5.5 kb fragment which would increase in intensity for each additional copy added. The results from the Southern blot analysis are shown in Fig. 5b. As expected, the M12 untransformed strain showed the 2.8 kb fragment, which corresponds to the intact *PGK1* locus. All transformed clones showed the 5.5 kb fragment thus confirming the *in tandem* integration of at least two copies of the vector. The observation that clone 5 showed other bands (including the 2.8 kb fragment) which were not present in the other clones suggests that, in this particular clone, the vector had integrated in a different manner. This is not entirely unexpected since it has recently been shown that a transforming cassette can integrate into the *K. phaffii*



genome in different cassette-to-cassette orientations and secondary recombination events may also occur [36]. Also, we cannot exclude the possibility of off-target integration events mediated by non-homologous end joining (NHEJ) which is the main repair system in filamentous fungi and higher eukaryotes [37]. Table 1 shows that faster growing clones 1, 4, 5 and 6 showed the highest vector copy number ( $\geq 14$  copies) as compared to slower growing clones 2 and 7 which had no more than 7 copies. Growth rate and copy number showed a linear correlation ( $R^2 = 0.8748$ ) (Fig. 5c). These results are in agreement with the prediction that faster growing clones would have more integrated copies of the defective marker.

### Genetic stability

Previous works have shown that single or low copy integrated messages are genetically stable in *K. phaffii* under different conditions [38, 39], however, few studies have focused on the integrity of multicopy clones. Since multicopy *K. phaffii* strains generally arise from multiple events of homologous recombination at the same locus, the integrated messages are typically repeated *in tandem*. The stability of such array may be compromised by excisional recombination which can “loop out” the genetic message under non-selective conditions [14], [40]. In order to test the stability of integrated pKGFP-I<sub>d</sub>, transformed cells were grown in non-selective medium (YPD) for 36 and 72 generations. The culture was transferred to fresh medium every 24 h to ensure that cells were in exponential phase throughout the experiment. After growth for 96 h or 144 h (36 and 72 generations, respectively), copy number of the selected clones was assessed by Southern blot analysis which showed that all clones maintained the original vector copy number (data not shown).

In a recent work, *S. cerevisiae* strains with multiple integrated cassettes bearing different defective auxotrophic markers also showed mitotic stability under prolonged nonselective conditions [41]. *S. cerevisiae* cells transformed with five or more copies of an integration vector conferring resistance to G418 and expressing *SUC2* (invertase) were very unstable during long-term culture in non-selective medium [42]. Likewise, when *K. phaffii* was transformed with a set of vectors containing sequentially increasing copies of porcine insulin precursor gene (PIP), both low and high copy strains were stable in serial culture in non-selective YPD medium. However, in high copy strains, loss of PIP cassettes was observed after 96 h of methanol induction [43].

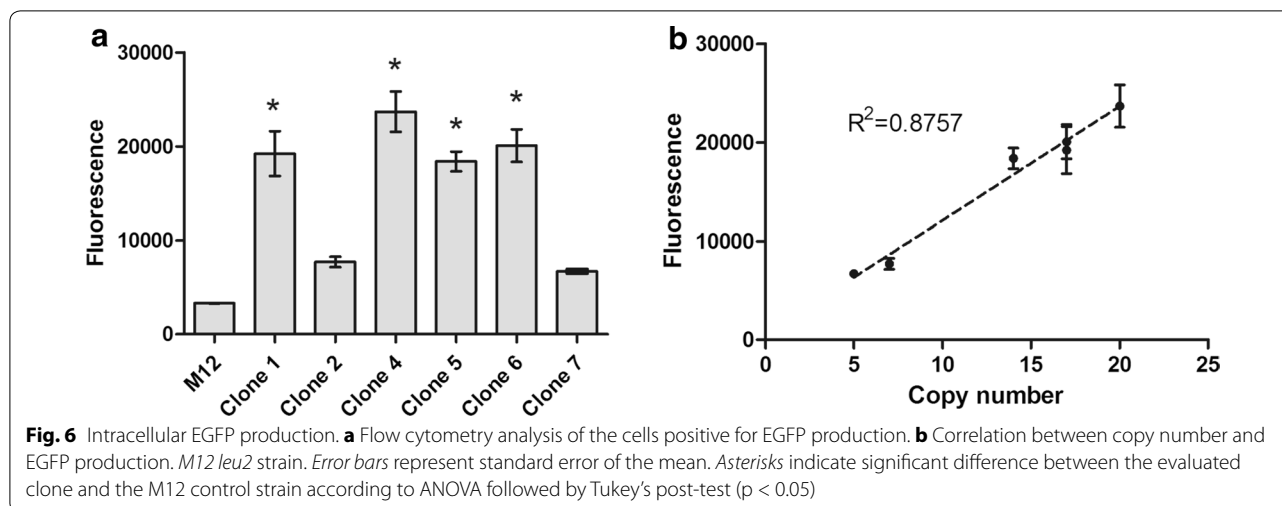
Based on these previous results, it has been proposed that multicopy strains should be carefully evaluated

for genetic stability especially under conditions of high expression or secretion [43]. In our work, since EGFP was produced intracellularly from a moderately strong *K. phaffii* promoter ( $P_{PGKI}$ ) [44], it is possible that the titers of this particular protein were not high enough to compromised cell growth as shown on Fig. 4, hence, genetic stability was observed.

### Correlation between copy number and protein production

In order to determine the correlation between plasmid copy number and increased heterologous protein production, intracellular fluorescence emission of each selected clone was determined by flow cytometry. As shown in Fig. 6a, all selected clones exhibited fluorescence emission higher than the untransformed M12 strain. EGFP production was the highest in clones 1, 4, 5 and 6 which exhibited higher plasmid copy number, whereas moderate production was observed in clones 2 and 7 (Table 1). Analysis of variance followed by Tukey's post-test showed significant difference in EGFP production in clones 1, 4, 5 and 6 when compared to M12 control strain ( $p < 0.05$ ). The EGFP fluorescence from clones 2 and 7 was not high enough to result in statistically significant differences in EGFP production in comparison to M12 ( $p > 0.05$ ). However, it is important to notice that the percentage of M12 cells producing EGFP (cells positive for EGFP) was less than 1%. As shown in Fig. 6b, we observed a linear correlation between vector copy number and EGFP production as measured by flow cytometry ( $R^2 = 0.8757$ ). It is expected that at a certain copy number the production of the heterologous protein might become detrimental to the cell; however, for the examined clones, we did not observe a decrease in cell viability nor in EGFP production which augmented linearly with up to 20 integrated copies. Similarly, an increase of up to eight copies of the hepatitis B surface antigen (HBsAg) gene showed a linear correlation with the concentration of mRNA and translated protein in *K. phaffii* [45].

However, the relationship between gene copy number and protein production is not always linear and in some cases it proved to be detrimental, especially for secreted proteins [6]. In another study involving EGFP, an increase of the secreted protein was observed with up to three copies but a decrease occurred with six copies [46]. When multicopy clones were used to produce intracellular human superoxide dismutase (hSOD) and secreted human serum albumin (HSA) a difference was observed in the correlation of gene copy number and productivity between non-secreted and secreted proteins [15]. The productivity of hSOD correlated linearly with gene copy number, while HSA productivity increased up to approximately 5–7 gene copies, and then decreased



with higher copy numbers. *K. phaffii* strains secreting human trypsinogen under the control of the *AOX1* promoter presented a positive correlation between copy number and product yield from 1 to 2 copies per cell, and a negative correlation at 3 or more copies [47]. Upon overexpression, great part of the heterologous protein was retained in the insoluble fraction of the endoplasmic reticulum. From this studies it is clear that bottlenecks in the secretory pathway are to some extent responsible for the low productivity of some multicopy clones [6].

Since the effect of gene dosage may vary from one protein to another, it is not possible to define the optimal copy number for any specific heterologous gene which should be assessed on a case-by-case basis. However, by using the approach presented in this work one can easily obtain a panel of clones with different copy numbers to be screened for the desired application. Furthermore, we envision that this approach might be also applied in synthetic biology studies in which different doses of specific genes may be required. This could be rapidly achieved by transforming M12 with different plasmids bearing the *leu2-d* marker following a screening for the desired phenotype. Work is underway in our laboratory to test this new application.

## Conclusions

In this work, we proposed a simple approach to obtain *K. phaffii* clones containing multiple copies of a desired expression vector. Our genetic system is based on a *K. phaffii* strain auxotrophic for leucine which is transformed with an expression vector bearing a defective *leu2-d* marker. The main advantage of the approach proposed here is the ease in selecting multicopy clones, in our case this was based on colony size. This approach might serve as a first step in the construction of strains

with higher productivity thus lowering the costs of industrial recombinant protein production.

## Methods

### Strains and growth conditions

*Komagataella phaffii* GS115 (*his4*) and X-33 (Invitrogen) were used as a source of template DNA to amplify *LEU2* and cell host to perform transformation with the disruption cassette, respectively. *K. phaffii* was routinely grown on YPD (1% yeast extract, 2% peptone and 2% glucose) at 28 °C. Solid medium was prepared by the addition of 2% agar. After transformation yeast cells were plated on YPD containing 300–500  $\mu\text{g mL}^{-1}$  G418 or 150  $\mu\text{g mL}^{-1}$  hygromycin B. Transformants were tested on MD [0.34% Yeast Nitrogen Base (YNB), 1% ammonium sulphate, 2% glucose, 0.4  $\mu\text{g mL}^{-1}$  biotin and 2% agar] and buffered MD [MD with 100 mM potassium phosphate (pH 6.0)] supplemented or not with 0.04 or 0.08% leucine. For induction of heterologous gene expression from the  $P_{AOX1}$  promoter cells were grown in a medium containing 1% yeast extract, 2% peptone, 100 mM potassium phosphate (pH 6.0), 0.34% YNB, 1% ammonium sulphate, 0.4  $\mu\text{g mL}^{-1}$  biotin supplied with 1% glycerol (BMGY medium) or 0.5% methanol (BMMY medium). When liquid medium was used, growth was carried out under agitation (200 rpm) in shake flasks with a volume at least 10 times greater than the volume of the medium.

Cloning procedures were carried out in *E. coli* XL10-gold (Stratagene, USA) which was cultivated in modified LB medium (0.5% yeast extract, 1% peptone and 1% NaCl) containing the appropriate antibiotic for selection of transformants (100  $\mu\text{g mL}^{-1}$  ampicillin or 50  $\mu\text{g mL}^{-1}$  kanamycin). Bacterial cells were grown at 37 °C with constant shaking (250 rpm). For solid medium, 1.5% agar was added.

## PCR

Phusion high-fidelity DNA polymerase (Finnzymes) was routinely used for PCR according to the instructions of the manufacturer. To amplify *LEU2*, Easy Taq DNA polymerase (LGC Bio, Brazil) was used in a final volume of 50  $\mu$ L consisting of 0.2 mM dNTP, 0.2  $\mu$ M each primer, 2 mM MgCl<sub>2</sub>, Easy Taq buffer 1X, 2 U polymerase and 1–5 ng template DNA. PCR involved an initial denaturation step at 96 °C for 3 min followed by 30 cycles of 60 s/94 °C, 1 min/60 °C, 2 min/72 °C and a final elongation step at 72 °C for 5 min.

## DNA manipulations

Plasmid extraction, electrophoretic analysis and other basic DNA manipulations were performed as described previously [48]. For DNA elution from agarose gels and for amplicon purification *Wizard SV Gel and PCR Clean-Up System* (Promega, USA) were used according to the manufacturer's instructions, respectively. Genomic DNA was purified by using *Wizard Genomic DNA Purification Kit* (Promega, USA) using the manufacturer's protocol.

## Construction of disruption cassette

First, *LEU2* was amplified by PCR from *K. phaffii* GS115 genome with primers PpLEU2-F1 and PpLEU2-R2 (Table 2) which introduce PvuII restriction sites at their 5'-ends. Primers were designed based on the published sequence of the *K. phaffii* chromosome 3 (accession # FR839630.1). The amplified 1.7 kb fragment included the *LEU2* coding region and ~300 bp of both upstream and downstream sequences. *LEU2* was cloned into pGEM-T easy (Promega) resulting in pGEM-LEU. This vector was digested with EcoRV to remove 375 bp of the coding region of *LEU2* (positions 405–750). To disrupt *LEU2*, first, a 1.7 kb fragment containing a *kan* expression cassette flanked by *loxP* sites was amplified from pPICK $\alpha$  [49] with primers ZeoBlas-F3 and ZeoBlas-R3

(Table 2). This amplicon was cloned into EcoRV-digested pGEM-LEU generating plasmid pLEU $\Delta$ kan. Finally, the *leu2::kan* disruption cassette was released after digestion of pLEU $\Delta$ kan with PvuII prior to yeast transformation.

## Marker excision

A vector based on pYRCre was constructed in order to promote marker excision in *K. phaffii*. Plasmid pYRCre was originally used to express the CreA recombinase in *S. cerevisiae* [49]. The P<sub>GALI</sub> promoter present in this vector was removed after XbaI digestion and replaced by a 441-bp fragment corresponding to the *S. cerevisiae* P<sub>TEFI</sub> promoter which was obtained by PCR using primers TEF-1F and TEF-1R (Table 2). The amplicon was digested with AvrII and cloned into XbaI-digested pYRCre. The resulting vector, pYRCre2, was used to transform *K. phaffii* and selection was made on YPD plates containing hygromycin B. Transformants were incubated at 28 °C for 3 days to allow expression of CreA recombinase and then selected clones were transferred to an YPD plate for plasmid curing. Isolated colonies were replica plated on YPD plates containing G418 or hygromycin B to confirm the removal of *kan* marker and cure of pYRCre2, respectively.

## Construction of expression plasmid pGFP-L2

First, a vector constructed in our lab derived from pPIC9 (Invitrogen, USA) with the EGFP reporter gene under the control of the P<sub>AOX1</sub> promoter was digested with EcoRV. This digestion removed the entire *HIS4* sequence, which was replaced by the *LEU2* gene obtained from pGEM-LEU after digestion with PvuII. The resulting vector, pPIC-LEU, was digested with BamHI and NotI to remove the EGFP gene which was fused in-frame to the  $\alpha$ -factor secretion signal. This secretable version of EGFP was replaced by a 741 bp EGFP fragment from pEGFP-N3 (Clontech, USA) after digestion of this plasmid with the same enzymes. The resulting plasmid, which allows

**Table 2 Primers used in this work**

Primer	Sequence (5'→3')*	Restriction site
PpLEU2-F1	<u>CAGCTGAAGAGTCCAAGTCCAAG</u>	PvuII
PpLEU2-R2	<u>CAGCTGGTGCCATTGGTGGTACTGT</u>	PvuII
TEF-2F	<u>ATACCTAGGCCCCACACCCATAGCTTCAA</u>	AvrII
TEF-2R	<u>ATACCTAGGTTTGTAATTAACCTTAGATTAGATTG</u>	AvrII
ZeoBlas-F3	<u>CGGATCCATAACTTCGTATAATGTATGCTATACGAAGTTATAGATCTCCACACCCATAGCTTCAAATG</u>	BamHI and BgIII
ZeoBlas-R3	<u>CGGATCCATAACTTCGTATAGCATACATTATACGAAGTTATAGATCTAGCTTGCAAATTAAGCCTTCGAG</u>	BamHI and BgIII
PpLEU2-EXT1	GAGGATAAGCTGGGAGACTATG	–
PpLEU2-EXT2	TCTGTTGCCTAAGACTGAGAGC	–
5-leud	GAGATCTATATATATTCAAGGATATACCATTCTAATG	BgIII
3-leud	GAGATCTGTTTCATGATTTTCTGTTACACC	BgIII

\* Restriction sites are underlined

intracellular expression of EGFP, was named pGFP-L2. Before *K. phaffii* transformation pGFP-L2 was linearized with *SacI* to promote targeted integration to the  $P_{AOXI}$  locus.

#### Construction of expression vector pKGFP-ld

The *leu2-d* allele was amplified by PCR from *S. cerevisiae* genome with primers 5- and 3-leud (Table 2). The amplified 1.4 kb fragment included the *LEU2* coding region with its transcription termination region and only 29 bp of its promoter [17]. The *leu2-d* amplicon was cloned into pBlueScript SK II (Agilent Technologies) and then liberated after *BglII* digestion for subcloning into *BamHI*-linearized pPICK2 [50] resulting in pK-ld vector. This vector was digested with *SacI* and *NotI* to remove the  $\alpha$ -factor secretory sequence. This digestion also removed a 179 bp fragment from  $P_{PGKI}$  which was restored when the EGFP gene was cloned. The 916 bp fragment including the EGFP gene fused to the 179 bp fragment from the  $P_{PGKI}$  was obtained by digestion of pPICK-GFP [a vector derived from pPIC9 (Invitrogen) for intracellular expression of EGFP under the control of  $P_{PGKI}$ ] with *SacI* and *NotI*. Cloning of this 916 bp fragment into pK-ld resulted in pKGFP-ld vector. This vector was linearized with *SacI* to target integration to the *PGKI* locus.

#### Yeast transformation

*Komagataella phaffii* X-33 was transformed by electroporation following the protocol described in the *Pichia* Expression Kit (Invitrogen). Transformation with pYRCre2 was carried out as previously described for the auto-replicative pPICHOLI vector [51].

#### Fluorescence microscopy

*Komagataella phaffii* cells expressing EGFP were grown in 5 mL BMGY for 16 h at 28 °C. After cell count, pelleted cells were re-suspended in 20 mL BMMY to a final OD<sub>600</sub> of 0.3. The culture was incubated at 28 °C and methanol was added to a final concentration of 0.5%. After 24 h of induction cells were imaged in a Zeiss Axio Observer Z1 Inverted Fluorescence Microscope equipped with 63× NA 1.4 oil immersion objective and a cooled CCD camera to analyze EGFP fluorescence. The images were acquired with Zen2011 software (Zeiss) and manipulated with Microsoft Office Picture Manager or Adobe Photoshop.

#### Growth kinetics

A fresh colony was inoculated in 500  $\mu$ L of MD medium in a *deep-well* plate and incubated for 24 h at 30 °C and 200 rpm. The appropriate volume of this culture was inoculated in 100  $\mu$ L of MD to an OD<sub>600</sub> = 0.08 in a

96-well plate. Cell growth was performed on the Epoch Microplate Spectrophotometer (Biotek) by incubating at 30 °C under agitation of 300 rpm for 72 h. OD<sub>600</sub> data was collected every 30 min. Three biological replicates were tested for each analyzed clone and the mean of the three values was presented. Natural logarithm of OD<sub>600</sub> values was used to construct growth curves. Maximal growth rate was calculated from the slope of the linear section of these curves (up to eight hours growth).

#### Southern blot analysis

Yeast cells were grown in 40 mL of MD medium at 30 °C under agitation during 24 h and the DNA was extracted using phenol–chloroform as previously described [48]. Approximately 10  $\mu$ g of genomic DNA were digested with *EcoRI* at 37 °C overnight. Digested DNA was applied in 0.8% agarose gel and then transferred to nitrocellulose membrane as described [48]. Probe labeling, hybridization and detection were made using the AlkPhos Direct Labeling and CDP-Star Detection System (GE Life Sciences) following specifications of the manufacturer. The probe used was a fragment of ~600 bp corresponding to the *PGKI* promoter obtained by digestion of pKGFP-ld with *BglII* and *BamHI*. The temperature for hybridization was 55 °C. Chemiluminescence was detected using the Amersham Imager 600 system (GE Life Sciences) and band intensity was measured with the use of the ImageQuant TL 8.1 software.

#### Genetic stability testing

The stability of the heterologous DNA integrated into the yeast genome was tested in shake flasks after 36 and 72 generations. A fresh colony was grown in 10 mL YPD medium for 24 h at 30 °C and 200 rpm. Then, 400  $\mu$ L of this pre-inoculum were inoculated in 40 mL YPD and incubated under the same conditions for 24 h. A 400  $\mu$ L sample of the culture was transferred to a new flask with 40 mL YPD and incubated under the same conditions for 24 h. This procedure was repeated four more times for a total growth time of 144 h. After 96 h (36 generations) and 144 h (72 generations) genomic DNA was extracted and submitted to Southern blot analysis as described above.

#### Flow cytometry

Yeast cells were grown in 5 mL of MD medium for 24 h at 30 °C and agitation. The required volume of each pre-inoculum was inoculated in 5 mL of MD to start the culture with an OD of 0.5. After 24 h of incubation at 30 °C under agitation cells were washed twice with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub> and 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) containing 0.5% Tween by centrifugation at 3000×g

for 5 min at 4 °C. Cells were suspended in the required volume of PBS to obtain approximately  $10^6$  cells mL<sup>-1</sup>. Cells were maintained at 4 °C until analysis with FACSVerse flow cytometer. All samples were collected with identical voltage parameters. Acquired data were analyzed using the FlowJo software. The gating strategy included: (a) gating on yeast cells on forward versus side scatter plots; (b) gating on single cells using forward scatter width versus forward scatter height plots and (c) selecting positive cells based on histograms from wild-type cells. Three biological replicates were tested for each analyzed clone and the mean of the three values is presented.

### Data analysis

Statistical analyses and figures were made on GraphPad Prims 5 software. ANOVA followed by Tukey's post-test was applied. Error bars on graphics represent standard error of the mean.

### Additional file

**Additional file 1.** Disruption of *LEU2* in *K. phaffii* X-33. Panel A: Annealing positions of primers used to amplify *LEU2* and for diagnostic PCR. Expected sizes of amplicons are in the bottom of each figure. The *LEU2* coding sequences are in dark grey. Primers: 1 PpLEU2-F1, 2 PpLEU2-R2, 3 PpLEU2-EXT1, 4 PpLEU2-EXT2, 5 ZeoBlas-F3. Panel B: PCR analysis in 1% agarose gel electrophoresis stained with ethidium bromide. M: O'GeneRuler 1 kb DNA ladder; lanes 1, 3 and 5: PpLEU2-EXT1/PpLEU2-EXT2; lanes 2, 4 and 6: ZeoBlas-F3/PpLEU2-EXT2. X-33: wild-type strain with intact *LEU2* gene, LK: *leu2* strain disrupted with *kan* cassette, M12: strain obtained after marker removal with CreA recombinase.

### Authors' contributions

MB carried out the experimental studies and drafted the manuscript. VR and JLDM participated in the design of the study and helped to draft the manuscript. AMN performed and analyzed the flow cytometry experiments. LM participated in the design of the study. FT conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

### Author details

<sup>1</sup> Laboratório de Biologia Molecular, Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF 70910-900, Brazil. <sup>2</sup> Faculdade de Medicina, Laboratório de Imunologia Celular, Universidade de Brasília, Brasília, DF 70910-900, Brazil.

### Acknowledgements

The authors wish to thank CNPq (Grant # 441978/2014-2) and FAPDF (Grant # 193.000.582/2009) for financial support of this Project.

### Competing interests

The authors declare that they have no competing interests.

### Availability of data and materials

All data and material concerning supporting the conclusions of this work is presented in the main paper and is made public available.

### Consent for publication

We consent BioMed Central to publish this manuscript should it be accepted for so.

### Funding

CNPq (Grant # 441978/2014-2), FAPDF (Grant # 193.000.582/2009).

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 17 April 2017 Accepted: 2 June 2017

Published online: 08 June 2017

### References

- Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M, Mattanovich D. *Pichia pastoris*: protein production host and model organism for biomedical research. *Future Microbiol.* 2013;8:191–208.
- Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol.* 2014;98:5301–17.
- Heyland J, Fu J, Blank LM, Schmid A. Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnol Bioeng.* 2010;107:357–68.
- Mellitzer A, Ruth C, Gustafsson C, Welch M, Birner-Grünberger R, Weis R, Purkardhofer T, Glieder A. Synergistic modular promoter and gene optimization to push cellulase secretion by *Pichia pastoris* beyond existing benchmarks. *J Biotechnol.* 2014;191:187–95.
- Lin Cereghino GP, Lin Cereghino J, Ilgen C, Cregg JM. Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr Opin Biotechnol.* 2002;13:329–32.
- Aw R, Polizzi KM. Can too many copies spoil the broth? *Microb Cell Fact.* 2013;12:128–36.
- Piva LC, Betancur MO, Reis VCB, De Marco JL, Moraes LMP, Torres FAG. Molecular strategies to increase the levels of heterologous transcripts in *Komagataella phaffii* for protein production. *Bioengineered* 2017;1–5. doi: [10.1080/21655979.2017.1296613](https://doi.org/10.1080/21655979.2017.1296613).
- Sreekrishna K, Kropp KE. *Pichia pastoris*. In: Wolf K, editor. *Nonconventional yeasts in biotechnology*. Berlin: Springer; 1996. p. 203–53.
- Wriessnegger T, Augustin P, Engleder M, Leitner E, Müller M, Kaluzna I, Schürmann M, Mink D, Zellnig G, Schwab H, Pichler H. Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of *Pichia pastoris*. *Metab Eng.* 2014;24:18–29.
- Romanos M, Scorer C, Sreekrishna K, Clare J. The generation of multicopy recombinant strains. *Methods Mol Biol.* 1998;103:55–72.
- Lin-Cereghino J, Hashimoto MD, Moy A, Castelo J, Orazem CC, Kuo P, Xiong S, Gandhi V, Hatae CT, Chan A, Lin-Cereghino GP. Direct selection of *Pichia pastoris* expression strains using new G418 resistance vectors. *Yeast.* 2008;25:293–9.
- Chen H-L, Yen C-C, Tsai T-C, Yu C-H, Liou Y-J, Lai Y-W, Wang M-L, Chen C-M. Production and characterization of human extracellular superoxide dismutase in the methylotrophic yeast *Pichia pastoris*. *J Agric Food Chem.* 2006;54:8041–7.
- Sunga AJ, Tolstorukov I, Cregg JM. Posttransformational vector amplification in the yeast *Pichia pastoris*. *FEMS Yeast Res.* 2008;8:870–6.
- Aw R, Polizzi KM. Liquid PTVA: a faster and cheaper alternative for generating multi-copy clones in *Pichia pastoris*. *Microb Cell Fact.* 2016;15:29–39.
- Marx H, Mecklenbräuker A, Gasser B, Sauer M, Mattanovich D. Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. *FEMS Yeast Res.* 2009;9:1260–70.
- Kazemi Seresht A, Nørgaard P, Palmqvist EA, Andersen AS, Olsson L. Modulating heterologous protein production in yeast: the applicability of truncated auxotrophic markers. *Appl Microbiol Biotechnol.* 2013;97:3939–48.
- Erhart E, Hollenberg CP. The presence of a defective *LEU2* gene on 2 mu DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number. *J Bacteriol.* 1983;156:625–35.

18. Hoylaerts M, Weyens A, Bollen A, Harford N, Cabezón T. High-level production and isolation of human recombinant alpha 1-proteinase inhibitor in yeast. *FEBS Lett.* 1986;204:83–7.
19. Serviené E, Melvydas V. Effect of the defective leucine gene *leu2-d*, on the properties of recombinant plasmid in yeast *Saccharomyces cerevisiae*. *Biologija.* 2001;4:30–3.
20. Ro D-K, Ouellet M, Paradise EM, Burd H, Eng D, Paddon CJ, Newman JD, Keasling JD. Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. *BMC Biotechnol.* 2008;8:83–96.
21. Lin Cereghino GP, Lin Cereghino J, Sunga AJ, Johnson MA, Lim M, Gleeson MA, Cregg JM. New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. *Gene.* 2001;263:159–69.
22. Nett JH, Gerngross TU. Cloning and disruption of the PpURA5 gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*. *Yeast.* 2003;20:1279–90.
23. Nett JH, Hodel N, Rausch S, Wildt S. Cloning and disruption of the *Pichia pastoris* ARG1, ARG2, ARG3, HIS1, HIS2, HIS5, HIS6 genes and their use as auxotrophic markers. *Yeast.* 2005;22:295–304.
24. Weninger A, Hatzl AM, Schmid C, Vogl T, Glieder A. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J Biotechnol.* 2016;235:139–49.
25. Bussey H, Umbarger HE. Biosynthesis of the branched-chain amino acids in yeast: a leucine-binding component and regulation of leucine uptake. *J Bacteriol.* 1970;103:277–85.
26. Wiame J, Grenson M, Arst HJ. Nitrogen catabolite repression in yeasts and filamentous fungi. *Adv Microb Physiol.* 1985;26:1–88.
27. Schreve J, Garrett JM. The branched-chain amino acid permease gene of *Saccharomyces cerevisiae*, *BAP2*, encodes the high-affinity leucine permease (S1). *Yeast.* 1997;13:435–9.
28. Grenson M, Hou C, Crabeel M. Multiplicity of the amino acid permeases in *Saccharomyces cerevisiae*. *J Bacteriol.* 1970;103:770–7.
29. Jauniaux J, Grenson M. GAP1, the general amino acid permease gene of *Saccharomyces cerevisiae*. Nucleotide sequence, protein similarity with the other bakers yeast amino acid permeases, and nitrogen catabolite repression. *Eur J Biochem.* 1990;190:39–44.
30. Kotliar N, Stella C, Ramos E, Mattoon J. L-leucine transport systems in *Saccharomyces cerevisiae* participation of GAP1, S1 and S2 transport systems. *Cell Mol Biol (Noisy-le-grand).* 1994;40:833–42.
31. Schreve JL, Garrett JM. Yeast Agp2p and Agp3p function as amino acid permeases in poor nutrient conditions. *Biochem Biophys Res Commun.* 2004;313:745–51.
32. Hueso G, Aparicio-Sanchis R, Montesinos C, Lorenz S, Murguía JR, Serrano R. A novel role for protein kinase Gcn2 in yeast tolerance to intracellular acid stress. *Biochem J.* 2012;441:255–64.
33. Horák J. Yeast nutrient transporters. *Biochim Biophys Acta.* 1997;1331:41–79.
34. Plant PJ, Manolson MF, Grinstein S, Demaurex N. Alternative mechanisms of vacuolar acidification in H<sup>+</sup>-ATPase-deficient yeast. *J Biol Chem.* 1999;274:37270–9.
35. Du M, Battles MB, Nett JH. A color-based stable multi-copy integrant selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as auxotrophic markers. *Bioeng Bugs.* 2012;3:32–7.
36. Schwarzzhans J-P, Wibberg D, Winkler A, Luttermann T, Kalinowski J, Friehs K. Integration event induced changes in recombinant protein productivity in *Pichia pastoris* discovered by whole genome sequencing and derived vector optimization. *Microb Cell Fact.* 2016;15:84.
37. Näätäsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS ONE.* 2012;7:e39720.
38. Ohi H, Okazaki N, Uno S, Miura M, Hiramatsu R. Chromosomal DNA patterns and gene stability of *Pichia pastoris*. *Yeast.* 1998;14:895–903.
39. Lim H-K, Kim K-Y, Lee K-J, Park D-H, Chung S-I, Jung K-H. Genetic stability of the integrated structural gene of guamerin in recombinant *Pichia pastoris*. *J Microbiol Biotechnol.* 2000;10:470–5.
40. Lee FW, Da Silva NA. Improved efficiency and stability of multiple cloned gene insertions at the delta sequences of *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol.* 1997;48:339–45.
41. Moon HK, Lee DW, Sim GH, Kim H-J, Hwang JY, Kwon M-G, Kang B-K, Kim JM, Kang HA. A new set of rDNA-NTS-based multiple integrative cassettes for the development of antibiotic-marker-free recombinant yeasts. *J Biotechnol.* 2016;233:190–9.
42. Wang X, Wang Z, Da Silva NA. G418 selection and stability of cloned genes integrated at chromosomal delta sequences of *Saccharomyces cerevisiae*. *Biotechnol Bioeng.* 1996;49:45–51.
43. Zhu T, Guo M, Sun C, Qian J, Zhuang Y, Chu J, Zhang S. A systematical investigation on the genetic stability of multi-copy *Pichia pastoris* strains. *Biotechnol Lett.* 2009;31:679–84.
44. Stadlmayr G, Mecklenbräucker A, Rothmüller M, Maurer M, Sauer M, Matanovich D, Gasser B. Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. *J Biotechnol.* 2010;150:519–29.
45. Vassileva A, Chugh D, Swaminathan S, Khanna N. Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*. *Protein Expr Purif.* 2001;21:71–80.
46. Liu H, Qin Y, Huang Y, Chen Y, Cong P, He Z. Direct evaluation of the effect of gene dosage on secretion of protein from yeast *Pichia pastoris* by expressing EGFP. *J Microbiol Biotechnol.* 2014;24:144–51.
47. Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D. Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnol Bioeng.* 2004;85:367–75.
48. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 1989.
49. Reis VCB, Nicola AM, de Souza Oliveira Neto O, Batista VDF, de Moraes LMP, Torres FAG. Genetic characterization and construction of an auxotrophic strain of *Saccharomyces cerevisiae* JP1, a Brazilian industrial yeast strain for bioethanol production. *J Ind Microbiol Biotechnol.* 2012;39:1673–83.
50. Arruda A, Reis VCB, Batista VDF, Daher BS, Piva LC, de Marco JL, de Moraes LMP, Torres FAG. A constitutive expression system for *Pichia pastoris* based on the *PGK1* promoter. *Biotechnol Lett.* 2016;38:509–17.
51. Lueking A, Horn S, Lehrach H, Cahill D. A dual-expression vector allowing expression in *E. coli* and *P. pastoris*, including new modifications. In: Vailancourt P, editor. *E. coli gene expression protocols SE—3*, vol. 205. Totowa: Humana Press; 2003. p. 31–42 (**Methods in Molecular Biology**<sup>™</sup>).

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

