



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

Construção de Cromossomo Artificial para *Komagataella phaffii*

Luíza Cesca Piva

Orientador: Prof. Fernando Araripe Gonçalves Torres

Coorientadora: Dr^a Viviane Castelo Branco Reis

Brasília, 2019

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

Construção de Cromossomo Artificial para *Komagataella phaffii*

Luíza Cesca Piva

Orientador: Prof. Fernando Araripe Gonçalves Torres

Coorientadora: Dr^a 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, 2019

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 e com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Banca Examinadora

Professora Dra. Cíntia Marques Coelho
Universidade de Brasília

Dr. Elibio L. Rech Filho
Embrapa Recursos Genéticos e Biotecnologia

Dr. João Ricardo M. de Almeida
Embrapa Agroenergia

*“Um navio no porto está seguro, mas
não é para isso que servem os navios.”*

Grace Hopper

Agradecimentos

Agradeço a Deus pela força.

Agradeço a todos do laboratório de Biologia Molecular que contribuíram para que este trabalho acontecesse.

Aos nossos técnicos Thompson e José, e à Márcia e Ivone pela amizade e muito trabalho.

Ao professor Fernando pela oportunidade e pela confiança.

À professora Janice pela amizade.

À prof^a Lídia pela disponibilidade em ajudar.

À Viviane pela coorientação, pelos conselhos e pela amizade.

A todos os colegas da equipe Lab 2: Marciano, Tulio, Yasmin, Igor, Dani, Pedro, Roberta, Carol, Tiago, Daniel, Diego, Chris, Catarina, Francisco, Vanessa, Hugo e Andressa.

Às “migas” Ana Laura, Myrna e Maritza, por todos os momentos de descontração. Por me apoiarem e me entenderem nos momentos mais difíceis deste doutorado.

À Júlia, minha estagiária, pela ajuda e maturidade impressionante.

Aos meus pais Maria Helena e Renir e minha irmã, Mariana, por tudo que sou hoje.

Ao meu marido Raphael, pelo companheirismo e apoio incondicional sem os quais essa jornada teria sido muito mais difícil.

A todos da EBB-UFPR pela minha formação.

Aos amigos de Curitiba e todos aqueles que moram longe, pela sintonia e carinho ao longo de tantos anos.

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

Ao CNPq pelo apoio financeiro.

Sumário

Lista de Figuras.....	i
Lista de tabelas	iii
Lista de abreviaturas e símbolos.....	iv
Resumo.....	v
Abstract.....	vi
1. Introdução	1
1.1. <i>Pichia pastoris</i> (<i>Komagataella phaffii</i>)	2
1.2. Ferramentas e estratégias de manipulação genética em <i>K. phaffii</i>	4
1.3. Cromossomos artificiais.....	7
1.4. Componentes do YAC de <i>K. phaffii</i>	9
2. Objetivos	12
2.1. Objetivo geral	12
2.2. Objetivos específicos.....	12
3. Estratégia.....	13
4. Material e Métodos.....	14
4.1. Linhagens e meios de cultura.....	14
4.2. Soluções.....	15
4.2.1. Soluções para preparo de células bacterianas competentes	15
4.2.2. Soluções para extração de plasmídeos por lise alcalina.....	15
4.2.3. Soluções para transformação de leveduras por eletroporação ...	16
4.2.4. Soluções para eletroforese em gel de poliacrilamida	16
4.2.5. Soluções para eletroforese de campo pulsado	17
4.2.6. Soluções para extração de DNA total de levedura.....	17
4.2.7. Soluções para transferência e Southern blot	17
4.2.8. Amplificação por PCR	18

4.3.	Manipulação do DNA e construção de plasmídeos	21
4.4.	Sequenciamento de plasmídeos	22
4.5.	Análises de bioinformática.....	22
4.6.	Preparação de células bacterianas competentes	23
4.6.1.	Preparo de células bacterianas quimicamente competentes	23
4.6.2.	Preparo de células bacterianas para eletroporação	24
4.7.	Extração de plasmídeos de <i>E. coli</i> por lise alcalina.....	24
4.8.	Mutagênese sítio-dirigida.....	25
4.9.	Transformação de leveduras por eletroporação	25
4.10.	Análise de estabilidade dos plasmídeos em meio líquido	27
4.11.	qPCR.....	27
4.12.	Eletroforese em gel de poliacrilamida não desnaturante	28
4.13.	Eletroforese de Campo Pulsado (PFGE).....	29
4.14.	Extração de DNA total de leveduras.....	30
4.15.	Southern Blot.....	31
5.	Resultados e discussão	32
5.1.	Capítulo 1: Construção de linhagens auxotróficas de <i>K. phaffii</i>	33
5.1.1.	Linhagem LA3	35
5.1.2.	Plasmídeo pPICH-Cre.....	38
5.1.3.	Linhagem LAU	41
5.2.	Capítulo 2: construção de vetores centroméricos.....	45
5.2.1.	Construção de vetor replicativo contendo gene <i>ADE3</i>	45
5.2.2.	Clonagem das sequências centroméricas.....	48
5.2.3.	Análise de estabilidade dos plasmídeos centroméricos.....	51
5.2.4.	Análise do número de cópias dos plasmídeos centroméricos.....	55
5.2.5.	Verificação da integridade dos vetores centroméricos	57
5.3.	Capítulo 3: construção do cromossomo artificial	62

5.3.1.	Amplificação e clonagem da sequência telomérica.....	63
5.3.2.	Análise da sequência telomérica clonada	66
5.3.3.	Preparo do vetor pBluetel-8	67
5.3.4.	Braço <i>ADE3</i>	71
5.3.5.	Braço CEN	72
5.3.6.	Transformação e montagem do YAC <i>in vivo</i>	78
5.3.7.	Verificação da integridade e estrutura dos YACs.....	81
6.	Conclusões e perspectivas	92
7.	Referências bibliográficas	94
8.	Anexos	11010

Lista de Figuras

Figura 1. Estratégia utilizada para a construção do cromossomo artificial de <i>K. phaffii</i> .	13
Figura 2. Via de biossíntese das purinas.	34
Figura 3. Deleção do gene <i>ADE3</i> .	36
Figura 4. Análise fenotípica da linhagem LA3.	37
Figura 5. Representação da amplificação do vetor pPICHOLI.	39
Figura 6. Confirmação da clonagem do cassete CreA no vetor pPICH.	40
Figura 7. Mapa do vetor pPICH-Cre.	41
Figura 8. Deleção do gene <i>URA5</i> .	42
Figura 9. Deleção do gene <i>LEU2</i> .	43
Figura 10. Confirmação da clonagem do cassete <i>ADE3</i> no vetor pPICH.	46
Figura 11. Determinação da orientação do gene <i>ADE3</i> .	47
Figura 12. Mapa do vetor pPICH-ADE3.	48
Figura 13. Anelamento dos <i>primers</i> para amplificação dos centrômeros de <i>K. phaffii</i> .	49
Figura 14. Confirmação da clonagem dos centrômeros.	50
Figura 15. Fenótipo da linhagem LA3 transformada com vetores centroméricos.	52
Figura 16. Teste de estabilidade mitótica.	53
Figura 17. Representação esquemática das regiões centroméricas de <i>K. phaffii</i> .	55
Figura 18. Determinação do número de cópias dos plasmídeos por célula.	56
Figura 19. Mapa do vetor pPICH-CEN3.	59
Figura 20. Southern Blot de DNA isolado da célula transformada com vetor pPICH-CEN3.	60
Figura 21. Confirmação da conformação do vetor pPICH-CEN4.	61
Figura 22. Representação do oligonucleotídeo TEL-PP.	63

Figura 23. Sítios de restrição de EcoRV e Afel usados na ligação das repetições teloméricas.	64
Figura 24. Estratégia da clonagem sequencial do fragmento contendo a repetição telomérica.	65
Figura 25. Confirmação da clonagem das repetições teloméricas.	66
Figura 26. Estratégia para construção do sítio de NdeI.	68
Figura 27. Confirmação da clonagem do fragmento contendo sequência NdeI.....	68
Figura 28. Estratégia de PCR e recombinação <i>in vitro</i> para adição de sítios de restrição ao vetor pBlue-tel8N.	69
Figura 29. Confirmação da clonagem do fragmento pKpn/pPOLI.	70
Figura 30. Confirmação da clonagem do cassete <i>ADE3</i> no vetor pBlue-tel8f.	71
Figura 31. Mapa do vetor pADE-tel8.	72
Figura 32. Confirmação da clonagem da sequência CEN4.....	73
Figura 33. Confirmação da clonagem do fragmento ARSzeo.	74
Figura 34. Mapa do vetor pCEN-tel8.....	75
Figura 35. Clonagem do fragmento P _{AOX1}	76
Figura 36. Mapa do vetor pCENA-tel8.	77
Figura 37. Representação esquemática da estratégia de montagem <i>in vivo</i> do cromossomo artificial.....	79
Figura 38. Confirmação da expressão de EGFP no clone YAC.	80
Figura 39. Diagrama representativo da construção do YAC final.....	82
Figura 40. PCRs de confirmação da montagem do vetor YAC.	83
Figura 41. PCRs de confirmação da montagem do vetor YAC-AOX1.....	84
Figura 42. Corrida curta de PFGE.....	86
Figura 43. Corrida longa de PFGE.	87
Figura 44. Southern Blot do clone YAC.....	90

Lista de tabelas

Tabela 1. Linhagens de <i>E. coli</i> e <i>K. phaffii</i> utilizadas neste trabalho.....	14
Tabela 2. Sequências 5'->3' dos <i>primers</i> utilizados neste trabalho.....	18
Tabela 3. Linhagens construídas neste trabalho.....	44
Tabela 4. Pares de <i>primers</i> e tamanho de fragmentos amplificados na verificação da conformação do vetor pPICH-CEN4.....	60
Tabela 5. Combinação de fragmentos lineares para transformação de <i>K.</i> <i>phaffii</i>	79
Tabela 6. <i>Primers</i> utilizados nas PCRs de confirmação dos clones YAC e YAC- AOX1.....	81

Lista de abreviaturas e símbolos

DNA	Ácido desoxirribonucleico
RNA	Ácido ribonucleico
<i>g</i>	Aceleração da gravidade
ARS	<i>Autonomously replicating sequence</i>
kb	Quilobase
pb	Par de bases
CEN	Centrômero
OD ₆₀₀	Densidade celular óptica medida a 600 nanômetros
PCR	Reação em cadeia da polimerase
rpm	Rotações por minuto
RNA-seq	Sequenciamento de RNA
YNB	<i>Yeast nitrogen base</i>
YPD	<i>Yeast peptone dextrose</i>
MD	<i>Minimal dextrose</i>
LB	Luria-Bertani
YAC	<i>Yeast artificial chromosome</i>
ORF	<i>Open reading frame</i>
EGFP	<i>Enhanced green fluorescent protein</i>
EDTA	Ácido etilenodiaminotetracético
IPTG	Isopropil-β-D-tiogalactopiranosídeo
X-gal	5-bromo-4-cloro-3-indolil-β-D-galactosídeo
DTT	Ditiotreitol
APS	Persulfato de sódio
MOPS	3-[N-morfolino]propano-ácido sulfônico
PFGE	<i>Pulsed-field gel electrophoresis</i>
SDS	Dodecil sulfato de sódio
NHEJ	<i>Non-homologous end joining</i>
IRES	<i>Internal ribosome entry site</i>
GMP	Guanosina monofosfato

Resumo

Komagataella phaffii é reconhecida como uma plataforma de produção de proteínas heterólogas, sendo capaz de crescer a altas densidades celulares e de produzir proteínas complexas. Novas ferramentas e construções têm permitido o uso desta levedura na produção de diversos tipos de moléculas, transformando-a em uma fonte potencial de aplicações biotecnológicas. Neste campo, as vias metabólicas do organismo são modificadas com a inserção, deleção ou manipulação de diversos genes para atingir o resultado desejado. Como obstáculos à aplicação desta levedura como plataforma biotecnológica, pode-se citar a instabilidade genética, a dificuldade de se trabalhar com grandes sequências de DNA e a necessidade de novas opções de sequências regulatórias para a manipulação de vias metabólicas. As soluções para estes problemas passam pelo desenvolvimento de ferramentas que permitam uma montagem mais rápida de grandes fragmentos de DNA a fim de inserir sequências maiores e mais complexas na levedura. Neste trabalho, foram desenvolvidas novas linhagens de *K. phaffii* contendo duplas auxotrofias: LA3, com auxotrofia para adenina e histidina, LAU, para adenina e uracila e LAL-kan, para adenina e leucina. Vetores contendo os centrômeros nativos da levedura foram construídos e avaliados quanto à segregação mitótica e ao número de cópias na célula. Ao final, um cromossomo artificial linear foi desenvolvido e testado na linhagem LA3. O conjunto de ferramentas moleculares desenvolvido nesse projeto representa um avanço para a manipulação genética de *K. phaffii*. Dessa forma, vias metabólicas e sistemas de regulação serão construídos mais eficiente e rapidamente, contribuindo para que a levedura alcance seu potencial como plataforma de produção de proteínas e metabólitos.

Abstract

Komagataella phaffii is recognized as a heterologous protein production platform, being able to grow at high cell densities and to produce complex proteins. New tools have allowed the yeast to be used on the production of a wide range of molecules and to become a potential source of biotechnological applications. In this field, metabolic pathways are modified through insertion, deletion or manipulation of many genes in order to achieve the desired result. Obstacles in the way of using this yeast as a biotechnological platform include genetic instabilities, difficulties when working with long DNA sequences and the need for new elements and parts for metabolic pathway regulation. The solutions for these problems involve developing tools that allow faster assembly of long DNA fragments in order to build bigger and more complex heterologous constructs. In this work, double auxotrophic strains were developed for *K. phaffii*: LA3, auxotrophic for adenine and histidine, LAU, for adenine and uracil, and LAL-kan, for adenine and leucine. We constructed centromeric vectors which were evaluated on the basis of their mitotic segregation and plasmid copy number per cell. Finally, a linear artificial chromosome was developed and tested in strain LA3. This new set of strains and vectors will represent an addition to the *K. phaffii* genetic toolbox and will contribute for the construction of whole metabolic pathways and allow the yeast to reach its full potential as a protein and metabolite production platform.

1. Introdução

O desenvolvimento da tecnologia do DNA recombinante, nos anos 70, deu início à produção de proteínas heterólogas através da manipulação genética de micro-organismos (ITAKURA *et al.*, 1977). Esta tecnologia inicialmente empregada em *Escherichia coli* logo foi desenvolvida para *S. cerevisiae*, permitindo a expressão de enzimas, hormônios, anticorpos e antígenos virais nessa plataforma, entre outras proteínas (BILL, 2012; CELIK; CALIK, 2012).

O desenvolvimento das tecnologias de sequenciamento e síntese de DNA aumentou as possibilidades de construção de vetores e de ferramentas para a manipulação genética de leveduras (MARDIS, 2017; TIAN; MA; SAAEM, 2009). Com o avanço das técnicas de biologia molecular, a quantidade de moléculas naturais e sintéticas produzidas por estes organismos aumentou exponencialmente. Leveduras foram utilizadas na produção de medicamentos, como a artemisina e derivados opióides, intermediários de indústrias químicas e de alimentos e biocombustíveis (DHAMANKAR; PRATHER, 2011; ESPAUX *et al.*, 2015; GALANIE *et al.*, 2015; RO *et al.*, 2006). Muitas destas aplicações deram origem a novas áreas de atuação, como a engenharia metabólica e a biologia sintética.

A engenharia metabólica consiste na manipulação genética de organismos a fim de modificar as funções enzimáticas, regulatórias e de transporte da célula. Esta disciplina trabalha com a mudança da configuração nativa do metabolismo celular, podendo utilizar enzimas e componentes já existentes e manipulá-los através da regulação dos genes ou inserir novas enzimas e rotas metabólicas no organismo (BAILEY, 1991). Já a biologia sintética compreende um conjunto de tecnologias que combina partes biológicas de maneira a construir genes, proteínas, vias metabólicas e/ou organismos completamente novos (CHURCH *et al.*, 2014). É uma área multidisciplinar, abrangendo a biologia computacional e molecular, a engenharia metabólica e a biologia de sistemas com o objetivo de criar ou conferir uma nova função a um organismo, desenhando circuitos com elementos que se combinam para atingir a resposta desejada (JENSEN;

KEASLING, 2014; PICATAGGIO, 2009). De certa forma, a biologia sintética contribui com a criação de novas ferramentas para a elaboração de estratégias de engenharia metabólica (LEE *et al.*, 2011).

Nesse contexto, *Komagataella phaffii* tem sido amplamente utilizada como plataforma de expressão heteróloga e, mais recentemente, de engenharia metabólica (SCHWARZHANS *et al.*, 2017b). Ferramentas de biologia sintética, como enzimas semissintéticas que modificam o padrão de glicosilação da levedura e promotores sintéticos de variadas intensidades já foram desenvolvidas e podem colaborar com a construção de linhagens cada vez mais sofisticadas e adaptadas às suas aplicações comerciais (HARTNER *et al.*, 2008; VOGL; HARTNER; GLIEDER, 2013).

1.1. *Pichia pastoris* (*Komagataella phaffii*)

P. pastoris é uma levedura haploide pertencente aos ascomicetos que foi isolada pela primeira vez em 1920 a partir de uma castanheira na França (GUILLIERMOND, 1920; CELIK; CALIK, 2012). É uma levedura metilotrófica cuja primeira aplicação comercial surgiu na década de 70, sendo considerada uma fonte de *single cell protein* para alimentação humana e animal. Este processo utilizava metanol como fonte de carbono e representava uma alternativa ao uso da soja como fonte de proteína nestes mercados (MACAULEY-PATRICK *et al.*, 2005). No entanto, devido ao aumento de preços derivado da crise do petróleo, este processo acabou tornando-se comercialmente inviável e a levedura perdeu importância comercial (CEREGHINO, CREGG, 2000).

Nos anos 80, foi observado que algumas de suas características a tornavam uma atrativa plataforma eucariótica para a produção de proteínas heterólogas, especialmente considerando o promotor forte e regulável do gene *AOX1* (VOGL; GLIEDER, 2013). *P. pastoris* apresenta vantagens como sua capacidade de atingir altas densidades celulares em cultivos líquidos e de realizar modificações pós-traducionais nas proteínas de interesse, com um padrão de glicosilação mais próximo ao humano que *S. cerevisiae* (AHMAD *et al.*, 2014; LI *et al.*, 2007). Desde então, centenas de proteínas (enzimas, peptídeos, proteínas bacterianas, virais e humanas) (CELIK; CALIK, 2012;

MACAULEY-PATRICK *et al.*, 2005) já foram produzidas nessa levedura, sendo algumas já comercializadas. Entre estas estão a insulina humana, vacinas contra hepatites B e C e anticorpos humanos (KIM *et al.*, 2015).

A produção de proteínas heterólogas neste sistema vem sendo complementada pela resolução de diversos gargalos do processo. Alguns aspectos considerados no melhoramento da levedura são os processos de dobramento e secreção de proteínas, alterações no balanço redox da célula, assim como a depleção de fatores intermediários do metabolismo e consequente queda nas taxas de crescimento (PEÑA *et al.*, 2018). Linhagens que apresentam um padrão de glicosilação de proteínas semelhante ao humano foram desenvolvidas para expandir a possibilidade de produção de proteínas com aplicações farmacêuticas (CHOI *et al.*, 2002). Assim, linhagens de *P. pastoris* com diversas modificações genéticas hoje proporcionam processos mais eficientes e produtos mais adequados às aplicações comerciais.

A engenharia metabólica de *P. pastoris* permitiu a produção de novas moléculas nesse sistema, como riboflavina, S-adenosil-metionina, licopeno, β -caroteno, ácidos orgânicos e ácido hialurônico de alto peso molecular (HE *et al.*, 2006; MARX; MATTANOVICH; SAUER, 2008; ARAYA-GARAY *et al.*, 2012; JEONG; SHIM; KIM, 2014; ZHANG *et al.*, 2015). *P. pastoris* possui potencial para a produção de metabólitos secundários graças ao forte fluxo metabólico na via das pentoses-fosfato e, sendo considerada uma levedura *crabtree-negativa*, seu metabolismo é fortemente aeróbico mesmo em grandes concentrações de substrato. Estas características favorecem a levedura como plataforma de produção e a diferenciam de *Saccharomyces cerevisiae* (SOLÀ *et al.*, 2004; POTVIN *et al.*, 2012).

O genoma de *P. pastoris* foi sequenciado pela primeira vez em 2009, sendo identificadas 5313 ORFs. A levedura possui 4 cromossomos, com 1,8, 2,2, 2,4 e 2,9 Mb. Hoje existem três linhagens já sequenciadas: GS115, DSMZ 70382 e CBS7435. Os dados do genoma estrutural e funcional podem ser encontrados no site www.pichiagenome.org (DE SCHUTTER *et al.*, 2009; KUBERL *et al.*, 2011). Posteriormente ao primeiro sequenciamento do genoma de *P. pastoris*, foram publicados diversos trabalhos refinando estes resultados e complementando a curadoria dos genes (STURMBERGER *et al.*, 2016;

VALLI *et al.*, 2016). A partir de estudos de transcriptoma, proteoma, metaboloma e secretoma desta levedura foi possível entender e manipular de maneira cada vez mais precisa seu metabolismo (MATTANOVICH *et al.*, 2009; CARNICER *et al.*, 2012; LOVE *et al.*, 2016).

Em 1995, a levedura *P. pastoris* foi reclassificada como pertencente ao gênero *Komagataella* e suas linhagens foram divididas em duas espécies. As linhagens CBS7435, X-33 e GS115 pertencem à espécie *Komagataella phaffii*, já as linhagens SMD protease-deficientes agora são conhecidas por *Komagataella pastoris* (YAMADA *et al.*, 1995; KURTZMAN, 2009; MATTANOVICH *et al.*, 2009). Neste trabalho, foram utilizadas linhagens correspondentes à espécie *K. phaffii*, portanto, a nomenclatura seguirá este padrão.

1.2. Ferramentas e estratégias de manipulação genética em *K. phaffii*

A manipulação genética de *K. phaffii* hoje é predominantemente realizada pelo uso de *shuttle vectors*, ou seja, vetores que são funcionais tanto em *Escherichia coli* quanto em *K. phaffii*. Muitos destes vetores estão comercialmente disponíveis como partes de sistemas de expressão que disponibilizam linhagens, vetores e técnicas para a expressão de genes de interesse de maneira constitutiva ou induzida (Pichia Expression Systems, Thermo Fischer Scientific).

A estratégia mais utilizada é a integração ao genoma, onde o plasmídeo contendo as sequências heterólogas é linearizado e tem sua integração orientada para sítios definidos, como os genes *HIS4* ou *AOX1* (AHMAD *et al.*, 2014). Esta integração ocorre através da recombinação homóloga das extremidades do vetor com o DNA genômico por adição, inserindo toda sua sequência no genoma da levedura. A recombinação também pode ocorrer de maneira a substituir uma região do genoma, utilizando cassetes de deleção de genes que recombinam em sequências espaçadas e excisam a região entre elas (KLINNER; SCHÄFER, 2004). A manipulação de *K. phaffii* por integração apresenta como vantagens a relativa estabilidade genética dos clones e a grande quantidade de vetores e informações disponíveis na literatura (MADSEN; VISMANS; BRINCH-PEDERSEN, 2016). Entretanto, muitas das

integrações ocorrem em locais inesperados do genoma, podendo interferir em processos metabólicos da célula (OBST; LU; SIEBER, 2017).

O uso de plasmídeos circulares capazes de se replicar na levedura, os chamados vetores replicativos, embora menos comum, apresenta maiores eficiências de transformação (LEE *et al.*, 2005). Estes vetores possuem sequências de replicação autônoma (ARS) que permitem sua manutenção na célula de maneira independente do genoma (CREGG *et al.*, 1985). Apesar de os plasmídeos replicativos apresentarem uma menor estabilidade mitótica ao longo das gerações de crescimento da levedura quando comparados a construções integradas no genoma, seu uso fornece um conjunto de clones mais homogêneo e permite a extração do vetor utilizado após a transformação, o que não é possível com estratégias de integração (SCHWARZHANS *et al.*, 2017).

Novas técnicas e tecnologias têm sido desenvolvidas com o objetivo de tornar mais prática e rápida a manipulação genética de *K. phaffii*. A montagem de sequências *in vivo* representa um significativo avanço no conjunto de ferramentas disponíveis, reduzindo o tempo necessário para obter construções complexas e aumentando o tamanho das sequências que podem ser inseridas em cada evento de transformação (YU *et al.*, 2012). Esta montagem pode ser realizada por ligação ou por recombinação, sendo que as sequências necessárias para uma recombinação bem-sucedida de fragmentos lineares de DNA são significativamente menores que as necessárias para uma integração genômica num *locus* específico (CAMATTARI *et al.*, 2016; SCHWARZHANS *et al.*, 2017). Enquanto a substituição de uma sequência genômica por uma estratégia integrativa pode demandar de 500 a 1000 pb de sequências homólogas para direcionar a recombinação, a montagem de vetores ocorre com sucesso com sequências de até 20 pb (CAMATTARI *et al.*, 2016; NÄÄTSAARI *et al.*, 2012; NETT; GERNGROSS, 2003). Uma vez que os vetores não precisam ser montados em *E. coli*, a limitação de tamanho imposta por esta plataforma de clonagem é removida e a complexidade e quantidade de elementos genéticos a ser inserida na levedura pode aumentar (CAMATTARI *et al.*, 2016).

Uma estratégia que busca facilitar a manipulação genética da levedura é a criação de linhagens adaptadas para uma eficiente recombinação homóloga.

Muitas das ferramentas moleculares usadas nesta levedura foram adaptadas de *S. cerevisiae*, que possui um sistema de reparo por recombinação homóloga bem definido. Resultados obtidos em *K. phaffii* mostram que a integração de DNA exógeno nesta levedura sofre interferência do sistema de recombinação não homóloga (NHEJ), causando uma série de integrações não específicas e reduzindo a eficiência da transformação. A integração em locais não desejados também é a responsável pelo conjunto heterogêneo de clones obtido em transformações que usam estratégias integrativas, pois dependendo do local de integração o metabolismo nativo pode ser afetado, assim como o nível de expressão dos genes heterólogos (SCHWARZHANS *et al.*, 2016).

A deleção do gene *KU70*, que codifica para uma proteína envolvida no complexo de reparo de extremidades de fita dupla de DNA, fez com que a parcela de clones com integrações aleatórias devido à NHEJ fosse reduzida. Entretanto, estas linhagens apresentaram instabilidades e baixas taxas de crescimento possivelmente causadas por estresses celulares e ainda não são recomendadas para uso industrial (NÄÄTSAARI *et al.*, 2012). Nesse sentido, o sistema CRISPR-Cas9 também busca contornar os problemas causados pela recombinação não homóloga e tem aumentado a eficiência tanto da inserção de mutações quanto das integrações direcionadas de sequências heterólogas no genoma da levedura (WENINGER *et al.*, 2016, 2017).

No contexto de construções utilizadas em engenharia metabólica, onde o tamanho total de múltiplos cassetes de expressão representa uma desvantagem na construção de plasmídeos, estratégias de expressão policistrônica oferecem uma maneira de otimizar o tamanho da sequência de DNA utilizada. A sequência viral 2A, que realiza a auto-hidrólise do peptídeo localizado entre proteínas, já foi utilizada na expressão de até 9 proteínas sob o controle de um mesmo promotor (GEIER *et al.*, 2015; WANG *et al.*, 2007). Outra estratégia utilizada na construção de cassetes policistrônicos é o uso de IRES (*Internal Ribosome Entry Site*), que permite a iniciação da tradução sem a necessidade de uma extremidade 5' de RNA mensageiro marcada pelo GMP metilado. Dessa forma, uma molécula de RNA pode ser traduzida para produzir mais de uma proteína. (LIANG *et al.*, 2012b; XU *et al.*, 2014).

1.3. Cromossomos artificiais

Os cromossomos artificiais de levedura (YAC) são vetores de clonagem de *S. cerevisiae* que se comportam como autênticos cromossomos no que tange à replicação e segregação mitótica. São capazes de receber fragmentos de DNA de até centenas de milhares de pares de bases (BURKE, CARLE, OLSON, 1985). Os YACs podem ter conformações circulares ou lineares, sendo que neste último caso necessitam de sequências que estabilizem suas extremidades livres (LARIONOV *et al.*, 1996; RECH *et al.*, 1990).

Para um vetor funcionar como um YAC linear são necessários uma sequência de replicação autônoma (ARS), centrômero e telômeros, todos elementos característicos de um cromossomo. Os elementos ARS são sequências que sinalizam o início da replicação do DNA e foram identificados em levedura pela capacidade de conferir replicação autônoma a plasmídeos (WILLIAMSON, 1985). Os telômeros são sequências ao redor das quais há um baixo nível de transcrição e que possuem a função de preservar as extremidades dos cromossomos e diferenciá-las de simples quebras na fita de DNA (WELLINGER, ZAKIAN, 2012). Em *S. cerevisiae*, foram isolados graças à sua capacidade de estabilizar vetores lineares (SZOSTAK; BLACKBURN, 1982). Já os centrômeros são responsáveis pela segregação independente de cromossomos durante a replicação celular, característica que permitiu seu isolamento pois estes elementos conferem estabilidade e um baixo número de cópias a plasmídeos replicativos de levedura (HIETER *et al.*, 1985b).

Os primeiros YACs foram desenvolvidos nos anos 80 e utilizados no estudo da segregação mitótica e meiótica, assim como na clonagem de sequências complexas (BURKE; CARLE; OLSON, 1985; MURRAY; SZOSTAK, 1983). Os telômeros utilizados nestas construções consistiam em sequências de DNA ribossomal da bactéria *Tetrahymena thermophila* e estimulavam a formação de telômeros nativos nas extremidades de vetores lineares (DANI; ZAKIAN, 1983; MONACO; ZOIA, 1994). Desde então estes vetores foram utilizados, entre outras aplicações, no estudo da estrutura e função de sequências de DNA de diversos organismos e na construção de bibliotecas para o sequenciamento do genoma humano (BRUSCHI; GJURACIC; TOSATO, 2012; MCPHERSON *et al.*, 2001).

Uma técnica importante utilizada neste tipo de vetor é a transformação e montagem *in vivo*, em apenas um passo, utilizando dois ou mais fragmentos de DNA. Esta técnica foi chamada de TAR (*Transformation-Associated Recombination*) e simplifica os procedimentos de clonagem quando comparada às técnicas convencionais (KOUPRINA *et al.*, 2003). Uma vez que a construção do vetor é realizada diretamente em células de levedura, YACs foram utilizados na montagem de sequências repetitivas de difícil clonagem e permitiram o isolamento de telômeros e centrômeros humanos, assim como a montagem dos primeiros cromossomos artificiais de mamíferos (IKENO *et al.*, 1998; RIETHMAN *et al.*, 1989). A montagem do primeiro genoma sintético, de *Mycoplasma genitalium*, foi feita primeiramente utilizando vetores de bactérias e posteriormente utilizando YACs quando os fragmentos sintéticos tornaram-se muito grandes (GIBSON *et al.*, 2008).

A aplicação de cromossomos artificiais pode ocorrer também na clonagem e expressão de diversas proteínas para a montagem de uma via metabólica. A via de síntese de flavonoides foi montada em *S. cerevisiae* com até 7 genes heterólogos utilizando uma plataforma chamada de eYAC (*expressible Yeast Artificial Chromosome*) (NAESBY *et al.*, 2009). Os cassetes de expressão de cada enzima foram concatenados de maneira aleatória e ligados aos braços de um cromossomo artificial linear. Como a capacidade de clonagem em YAC é de centenas de pares de bases, podem-se clonar várias cópias dos genes envolvidos na via e selecionar pela maior produção do metabólito de interesse (ESSANI; GLIEDER; GEIER, 2015). Cromossomos artificiais já foram utilizados também na construção de bibliotecas combinatórias de cDNA para a produção de moléculas de interesse para a indústria farmacêutica (KLEIN *et al.*, 2014).

Apesar de não terem a mesma função dos cromossomos artificiais, os cromossomos sintéticos de *S. cerevisiae* têm sido utilizados para o estudo das sequências mínimas essenciais para o funcionamento da célula, assim como do efeito de diversas deleções e mutações (DYMOND *et al.*, 2011). O projeto Sc2.0 reuniu diversos grupos de pesquisa para criar o primeiro genoma eucariótico sintético e utilizou os cromossomos nativos como estruturas nas quais sequências sintéticas eram inseridas por recombinação homóloga e substituíam as sequências originais (RICHARDSON *et al.*, 2017). Além disso,

foram construídas linhagens de *S. cerevisiae* com dois ou até apenas um cromossomo circular ao invés dos 16 cromossomos encontrados nas linhagens selvagens. Estas moléculas tiveram seus telômeros e grande parte de seus centrômeros excisados e suas extremidades ligadas, formando um ou dois grandes cromossomos (LUO *et al.*, 2018; SHAO *et al.*, 2018). Estudos como estes têm como objetivo estudar a evolução de eucariotos e podem contribuir para um maior entendimento da manutenção de telômeros, da replicação de moléculas circulares e da estrutura tridimensional dos centrômeros.

1.4. Componentes do YAC de *K. phaffii*

K. phaffii ainda não possui cromossomos artificiais para uso como vetores de clonagem ou expressão, porém as sequências essenciais para a construção de YACs já foram identificadas. Os centrômeros de cada um dos quatro cromossomos de *K. phaffii* foram localizados a partir de análises experimentais e *in silico*. Visto que a estrutura tridimensional dos centrômeros é um componente muitas vezes mais importante para sua função que a própria sequência de DNA, a busca utilizou uma proteína nativa homóloga à proteína centromérica CSE4. Esta proteína é uma histona que faz parte da cromatina do centrômero, que por sua vez liga os cromossomos duplicados ao fuso mitótico durante a mitose e a meiose (COUGHLAN *et al.*, 2016). A proteína CSE4 foi marcada de maneira fluorescente e revelou um padrão de organização centromérica típico de leveduras (STURMBERGER *et al.*, 2016). A análise tridimensional da conformação do genoma de *K. phaffii* também indicou potenciais regiões que poderiam conter cada um dos quatro centrômeros, em janelas de 20 kb (VAROQUAUX *et al.*, 2015). Além disso, uma extensa análise de RNA-seq sequenciou e quantificou os transcritos desta levedura e indicou quatro regiões correspondentes a uma baixa taxa de transcrição. Como centrômeros possuem uma complexa estrutura tridimensional e geralmente não possuem genes expressos em sua extensão, estas regiões, que eram compatíveis com aquelas obtidas pelos outros métodos, corresponderiam aos centrômeros de *K. phaffii* (LIANG *et al.*, 2012a; LYN CHAN; WONG, 2012; STURMBERGER *et al.*, 2016). Finalmente, foi verificado que estas regiões consistem em repetições invertidas de sequências de aproximadamente 2,5 kb,

separadas por um segmento central de 800 a 1300 pb e não possuem similaridade de sequência com centrômeros de nenhuma outra levedura (COUGHLAN *et al.*, 2016).

Sequências de replicação autônoma já vêm sendo usadas em vetores de *K. phaffii* há mais de 30 anos (CREGG *et al.*, 1985). A sequência ARS1 localiza-se no cromossomo 2 da levedura e está disponível comercialmente no vetor pPICHOLI (MoBiTec) (CREGG *et al.*, 1985). Mais recentemente, novas sequências foram testadas, com melhor desempenho quanto à estabilidade do vetor (LIACHKO *et al.*, 2014). Uma sequência de replicação autônoma de origem mitocondrial foi isolada recentemente e vetores replicativos têm sido construídos para aplicações da levedura em biologia sintética (CAMATTARI *et al.*, 2016; SCHWARZHANS *et al.*, 2017). Outro aspecto importante quanto às ARS de *K. phaffii* é a presença de sequências deste tipo dentro e ao redor dos centrômeros dos cromossomos 2, 3 e 4 (COUGHLAN *et al.*, 2016; LIACHKO *et al.*, 2014).

Os telômeros de *K. phaffii* foram identificados a partir do sequenciamento completo de seu genoma, sendo compostos por repetições em *tandem* da sequência TGGATGC. Estas repetições variam de 100 a 340 pb e se assemelham às sequências teloméricas de *Yarrowia lipolytica* e *Ogataea polymorpha*, mas não àquelas encontradas em *S. cerevisiae* (HANSON; BYRNE; WOLFE, 2014; LOVE *et al.*, 2016; STURMBERGER *et al.*, 2016). A manutenção das sequências teloméricas é geralmente realizada por enzimas como a telomerase, responsável pela replicação de extremidades de DNA (WELLINGER; ZAKIAN, 2012). Em *K. phaffii*, não há muitos estudos elucidando o mecanismo de manutenção dos telômeros, porém sabe-se que o dímero proteico Ku70/Ku80 participa deste processo em menor escala que em *S. cerevisiae* (NÄÄTSAARI *et al.*, 2012). Em posição adjacente ao telômero do cromossomo 4 da levedura, encontram-se dois genes *MAT* que participam da inversão por recombinação que ocorre na troca de *mating type* (HANSON, BYRNE, WOLFE, 2014).

Além destes componentes básicos, a construção de cromossomos artificiais lineares envolve duas marcas de seleção porque a transformação é realizada com dois braços lineares e um inserto (SLEISTER *et al.*, 1992).

Assim, cada um dos braços de seleção tem uma marca diferente para seleção da correta montagem do YAC.

Considerando que todos os componentes de um cromossomo artificial já foram identificados em *K. phaffii*, a construção de um vetor deste tipo uniria estas recentes descobertas e permitiria a clonagem de grandes sequências, facilitando o uso da levedura como plataforma não apenas de expressão heteróloga, mas também de engenharia metabólica. Construções de DNA usadas em engenharia metabólica e biologia sintética podem incluir diversas enzimas, sítios de ligação e proteínas regulatórias, fatores de transcrição e polimerases sintéticas, compondo grandes extensões de DNA (NIELSEN *et al.*, 2013; PINHEIRO *et al.*, 2012).

Dessa forma, em vista da necessidade de ferramentas que tornem a manipulação genética de *P. pastoris* mais prática e rápida, os cromossomos artificiais se mostram como alternativas adequadas já que são capazes de receber sequências muito maiores que os vetores comuns de levedura. A capacidade de realizar a transformação associada à recombinação, eliminando a passagem das construções por *E. coli*, também amplia a variedade de construções disponíveis. O emprego de um YAC em *K. phaffii* aumentaria o potencial deste organismo como plataforma de produção de moléculas de interesse biotecnológico através de tecnologias baseadas em engenharia metabólica e biologia sintética.

2. Objetivos

2.1. Objetivo geral

Construir um cromossomo artificial linear para *K. phaffii*.

2.2. Objetivos específicos

- a) Obter linhagens de *K. phaffii* com dupla auxotrofia (*ade2 ade3*; *ade2 leu2* e *ade2 ura5*);
- b) Construir vetor replicativo com a marca auxotrófica *ADE3*;
- c) Construir vetor centromérico e analisar a estabilidade dos quatro centrômeros de *K. phaffii*;
- d) Construir o vetor YAC linear;
- e) Avaliar estrutura e função do cromossomo artificial.

3. Estratégia

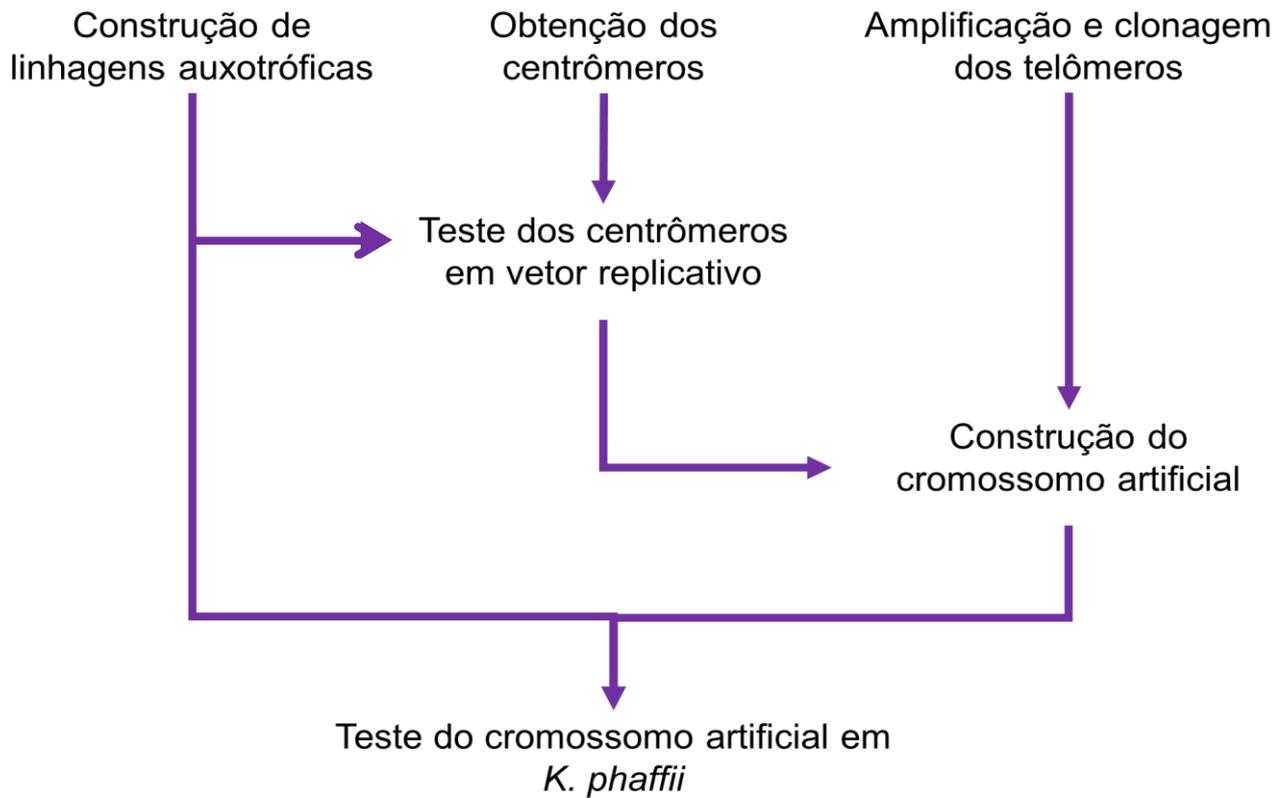


Figura 1. Estratégia utilizada para a construção do cromossomo artificial de *K. phaffii*.

4. Material e Métodos

4.1. Linhagens e meios de cultura

As linhagens de microrganismos utilizadas e construídas neste trabalho estão listadas na tabela 1.

Tabela 1. Linhagens de *E. coli* e *K. phaffii* utilizadas neste trabalho.

Linhagem	Genótipo	Fonte
<i>E. coli</i> XL10-gold	endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet ^R F'[proAB lacI ^q ZΔM15 Tn10(Tet ^R Amy Cm ^R)]	Agilent Technologies
<i>E. coli</i> DH10B	str. K-12 F ⁻ Δ(ara-leu)7697[Δ(rapA'-cra')] Δ(lac)X74[Δ('yahH-mhpE) duplication(514341-627601)[nmpC-gltI] galK16 galE15 e14 ⁻ (icd ^{WT} mcrA) ϕ80dlacZΔM15 recA1 relA1 endA1 Tn10.10 nupGrpsL150(Str ^R) rph ⁺ spoT1 Δ(mrr-hsdRMS-mcrBC) λ ⁻ Missense(dnaA glmS glyQ lpxK mreC murA) Nonsense(chiA gatZ fhuA? yigA ygcG) Frameshift(flhC mglA fruB)	Thermo Fischer Scientific
<i>E. coli</i> BMH 71-18	thi, supE, Δ(lac-proAB), [mutS::Tn10], [F', proAB, lacI ^q ZΔM15]	Promega
<i>K. phaffii</i> X-33	Wild-type	Thermo Fischer Scientific
<i>K. phaffii</i> LA2	amd2 ade2	(PIVA <i>et al.</i> , 2018)

As leveduras foram cultivadas em meio complexo YPD (1% extrato de levedura, 2% peptona e 2% glicose) e meio mínimo MD (0,34% YNB [Difco],

4×10^{-5} % biotina, 2% glicose e 1% sulfato de amônio). No caso de meio sólido foi adicionado 2% ágar. Para as leveduras transformadas com vetores contendo zeocina como marca de seleção foi usado o meio YPDS (1M sorbitol foi adicionado ao meio YPD ágar). Quando necessário, foi adicionado 100 $\mu\text{g/mL}$ zeocina, 200 $\mu\text{g/mL}$ higromicina ou 500 $\mu\text{g/mL}$ G418 ao meio complexo. Na seleção de clones contendo a marca de seleção *amdS* (meio MD-Ac) o sulfato de amônio foi substituído por 0,6% acetamida como fonte de nitrogênio no meio MD. Nos crescimentos para contrasseleção foi adicionado ao meio MD 0,23% fluoroacetamida ou 0,1% 5-FOA. Quando necessário, o meio MD foi suplementado com 0,002% adenina, 0,001% uracila, 0,08% leucina e 0,004% histidina.

O cultivo de bactérias foi feito em meio LB (0,5% extrato de levedura, 1% peptona e 1% NaCl. Foi adicionado ágar 1,5% no caso de meio sólido). Para vetores com a marca de seleção para zeocina, foi usado o meio LB *low-salt* com apenas 0,5% NaCl. Quando necessário, foram adicionados 25 $\mu\text{g/mL}$ zeocina, 50 $\mu\text{g/mL}$ kanamicina e 100 $\mu\text{g/mL}$ ampicilina. Quando da seleção azul e branco em clonagens, 0,1 mM IPTG e 0,004% X-gal foram adicionados ao meio LB sólido. Para o preparo de células competentes, foi utilizado o meio SOB (2% triptona, 0,5% extrato de levedura, 0,06% NaCl, 0,05% KCl e 10 mM de cada um dos sais: MgCl_2 e MgSO_4).

4.2. Soluções

4.2.1. Soluções para preparo de células bacterianas competentes

- Tampão de transformação I: 12 g/L RbCl; 9,9 g/L $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$; 0,03 M acetato de potássio; 1,5 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ e 150 g/L glicerol.
- Tampão de transformação II: 0,02 M MOPS; 1,2 g/L RbCl; 11 g/L CaCl_2 e 150 g/L glicerol.

4.2.2. Soluções para extração de plasmídeos por lise alcalina

- TE: 25 mM Tris-HCl; 10 mM EDTA (pH 8,0).

- Solução II: 0,2 M NaOH; 1% SDS.
- Solução III: 3 M acetato de sódio; 2 M ácido acético (pH 5,0).

4.2.3. Soluções para transformação de leveduras por eletroporação

- Solução de pré-tratamento: 100 mM acetato de lítio; 10 mM DTT; 0,6 M sorbitol; 10 mM Tris-HCl (pH 7,5).

4.2.4. Soluções para eletroforese em gel de poliacrilamida

- Acrilamida:bisacrilamida: 29% acrilamida; 1% bisacrilamida.
- 1,5 M Tris-HCl (pH 8,8).
- 1 M Tris-HCl (pH 6,8).
- 10% SDS.
- 10% Persulfato de amônio.
- Tampão de amostra não desnaturante 5X: 60 mM Tris-HCl (pH 6,8); 25% glicerol; 0,1% azul de bromofenol.
- Tampão de corrida Tris-Glicina 5X: 16,7 g/L Tris; 104,5 g/L glicina; 0,5% SDS.
- Solução de lise: 50 mM Tris-HCl (pH 8,0); 1,25% Tween 20; 13,5 µL de mix de inibidores de proteases (solução estoque do mix de inibidores: 1 tablete de *Complete Tablets Mini EDTA-free Easy pack* dissolvida em 1 mL de água destilada)

4.2.5. Soluções para eletroforese de campo pulsado

- Solução de esferoplastos: 7,5% β -mercaptoetanol; 0,5 M EDTA (pH 8,0); 0,1 mg/mL liticase.
- Solução de protease: 1,5% sarcosil; 0,5 M EDTA (pH 8,0); 1 mg/mL proteinase K.
- TAE 1x: 40 mM Tris; 20 mM ácido acético; 1 mM EDTA.
- TBE 0,5x: 45 mM Tris; 45 mM ácido bórico; 1 mM EDTA.

4.2.6. Soluções para extração de DNA total de levedura

- Tampão SE: 0,9 M sorbitol; 100 mM EDTA (pH 7,5).
- Tampão TE20: 10 mM tris-HCl pH 7,5; 20 mM EDTA (pH 7,5).

4.2.7. Soluções para transferência e Southern blot

- Solução de depurinação: 0,25 M HCl.
- Solução de desnaturação: 0,5 M NaOH; 1,5 M NaCl.
- Solução de neutralização: 1 M acetato de amônio; 0,02 M NaOH.
- SSC 20X: 3 M citrato de sódio; 5 M NaCl.

4.2.8. Amplificação por PCR

A amplificação de fragmentos por PCR foi realizada com diferentes enzimas, dependendo da aplicação. Foram utilizadas a Phusion DNA Polymerase (Thermo Scientific), enzima de alta fidelidade, a Taq DNA Polymerase (Invitrogen), MyTaq DNA Polymerase (Bioline) e a GoTaq G2 DNA Polymerase (Promega) para PCRs de confirmação e a Accutaq LA DNA Polymerase (Sigma Aldrich) para amplificar fragmentos longos. No caso de PCR de colônia, uma pequena porção de células foi adicionada ao tubo de reação com um palito estéril e a etapa de desnaturação inicial teve um acréscimo de 5 minutos. Todas usadas de acordo com as instruções dos fabricantes. Os *primers* utilizados neste trabalho estão na tabela 2.

Tabela 2. Sequências dos *primers* utilizados neste trabalho. Sítios de restrição estão sublinhados. Em minúsculo, sequências para recombinação.

Primer	Sequência (5'→3')	Sítios de restrição
ADE3up-F	gataagcttgatatcgaattcctgcag <u>ccccccggg</u> ACGTAATG GAATAACTGCTGAC	Smal
ADE3up-R	gaagttatggatcctACGAGGTAATTGAAGGCTCAC	
ADE3lox-F	cttcaattacctcgtAGGATCCATAACTTCGTATAATG	
ADE3lox-R	caatctctccctgtcatCGGATCCATAACTTCGTATAG	
ADE3dw-F	gaagttatggatccgATGACAAGGGAGAGATTGAAG	
ADE3dw-R	ggtggcggccgctctagaactagtggat <u>ccccccggg</u> TGCAAT GTA CTGTTGAGTAGG	Smal
ADE3conF	GGGGACCGGAGGTA AAAAGAC	
ADE3conR	GTTGGAATAATTGCATGGTCTG	
URA5F	CGCTCTTGCCTCTGCTGATGTTC	
URA5R	ATCTTGAGCCTGAGTGTCTTCA	
PpLEU2-EXT1	GAGGATAAGCTGGGAGACTATG	
PpLEU2-EXT2	TCTGTTGCCTAAGACTGAGAGC	
pARS1	AAACCTAGGCAATTAATATTTACTTATTTTGGTCA ACCCCAAATAGGTTGATTTCATACTTGGTTCATT	

	CAAAAATAAGTAGTCTTTTGAGATCTTTCAATATT ATAATAAATATACTATAACAGCCGACTTGTTTCA TTTTCGCGAATGTTCCCCAGCTTATCACTAGTC CCCACACACCATAGCTTCAAATG	
pPIC-F	GATTCAGAATTCGTTTAAACGCGGCCGCCCTAG GCAATTAATATTTACTTATTTTGG	EcoRI PmeI NotI
pPIC-R	GATTCAGAATTCGGATCCGTTTAAACTCATTCAA CCTTTCGTCTTTGG	EcoRI BamHI PmeI
MUT1(Hpa)	CCATTGACATGGTAAACAGTTGGA	
MUT2(Bam)	GAACACTGGGATCTTGGTTGAGG	
CEN1-F	gagtttaaaccgcatccACGAAGCAATGGATAGGCACT	BamHI
CEN1-R	ccgcgaattcggatccTGAAGTCTTTCAGAGAGGAGCA	BamHI
CEN1c-F	CAAGTATGCGTGATCCCAGGT	
CEN1c-R	TACGAATTGTGGGGCTCTGT	
CEN2-F	gagtttaaaccgcatccATCTCCGTTGATACTCCCAAC	BamHI
CEN2-R	ccgcgaattcggatccATCGACAAGCAGAACACTAAG	BamHI
CEN2c-F	GAATGGAGGTGCTGGTGGTTA	
CEN2c-R	TGTAATGCTCGCTGGTGAGT	
CEN3-F	gagtttaaaccgcatccAAGTGGTACACCAGTCAGCG	BamHI
CEN3-R	ccgcgaattcggatccTCAGTATTCAACTGCAACTGC	BamHI
CEN3c-F	TCAGCCGAATACCCACACTT	
CEN3c-R	TCAGCCGTCAGCGAAATGAT	
CEN4-F	gagtttaaaccgcatccCAAACGCACCGTCTTGTTCA	BamHI
CEN4-R	ccgcgaattcggatccAATTGATGTAGACGAGCAGC	BamHI
CEN4c-F	TCAAGAATCGTACTGGCACCT	
CEN4c-R	CAAGCTCGTGAGATGGGATGT	
CEN370-F	Cgctcagtggaacgaaaactcacgttaagggatttggtcatgagatc agatctaacatccaaagacgaaaggtgaatgagtttaaaccgcatcc aagtg	
CEN370-R	Gctggccctctctcccagctcacgaatcagatcctaagtcctactcaa cagtacattgcagcggccgcggttaaaccgaattcggatcctcagtattc	

	aa	
TEL-PP	ACTCACtatagggcgaattgggtaccAGCGCTGGATGCT GGATGCTGGATGCTGGATGCTGGATGCTGGAT GCTGGATGCTGGATGCTGGATGCTGGATGCTG GATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATAT <u>CgagctccagcttttgttcccTTTAGT</u>	KpnI AfeI EcoRV SacI
OTEL-F	TATAGGGCGAATTGGGTACCAG	KpnI
OTEL-R	GGGAACAAAAGCTGGAGCTCGA	SacI
M13-F	GTAAAACGACGGCCAG	
M13-R	CAGGAAACAGCTATGAC	
qZEO-F	CGACGTGACCCTGTTCATCA	
qZEO-R	TGGACACGACCTCCGACCA	
qHIS-F	GTGTATCCTGGCTTGGCATCT	
qHIS-R	GCCAAGTACGGTGTGACGTT	
5zeo	CGGATCCATGGCCAAGTTGACCAGT	
3zeo	CGGATCCTCAGTCCTGCTCCTCGG	
a-s-F	ACTGGTTCGCCACTCTTGTT	
a-s-R	TCAGACAGCCTTCAATGGGC	
pKpn	agcatccagcaatccagcgctggtacCCAATTCGCCCTATA GTGAGTCG	KpnI
pPOLI	agagcttgacggggaaagccggcCATATGGCGGCCGCC CTAGGGGATCCGAACGTGGCGAGAAAGGAAG	NotI AvrII BamHI NdeI
ARSzeo-F	ttgaacaagacggtgctgttgggatcccctagggcggccgcCAATT AATATTTACTTATTTTGG	
ARSzeo-R	ccgatttagagcttgacggggaaagccggccatatggcggccgcTT GCTCACATGTTGGTCTCC	
AOX1-F	ccaaaataagtaaataattaattggcggccgcctaggCTAACATC CAAAGACGAAAGGTTG	
AOX1-R	ttgaacaagacggtgctgttgggatcccctaggTCCTCGTTTTCG AATAATTAGTTG	

amdS-GFP	atcattcgagtttgatgttttcttggtatttcccactcctcttcagagtacag aagattaagtgagaccCCCACACACCATAGCTTCAAAA	
amdS-ARS	tactgaaaaccttgcttgagaaggtttgggacgctcgaaggcttaattt gcaagctggagaccaacatgtgagcaagcggccgcatatgGCA AATTAAGCCTTCGAGCGTC	
GFP-ADE3	tcttctgcgagtttctcatttgtgaatgggcatgtgttccaatactttatta tttggcagcagttattccattacgtgcggccgcatatgGCGAGGC AAGCATCTACTAAT	
GFP-amdS	tgcgcgagtgccgagaaaaatctggaagagtaaaaaaggagtagaa acatttgaagctatggtgtgtgggGGTCTCACTTAATCTTC TGTAATC	
c-ADEGFP-F	ACTTCATCAGCGTAGTTAGCCA	
c-ADEGFP-R	TACCACAAGAGGTGGCATCAC	
c-amdSGFP-F	ACATGACTGTTCCCTCAGTTCAAG	
c-amdSGFP-R	ATGGAGGTCACTGAAAGAG	
qAMS-F	CTGGACACACCGTCACACCTT	
qAMS-R	GATGTTTGGAAATGGCAGGCTCTC	

4.3. Manipulação do DNA e construção de plasmídeos

A construção dos vetores utilizados neste trabalho utilizou técnicas básicas de biologia molecular descritas por Sambrook e colaboradores (2001). As digestões de fragmentos de DNA com enzimas de restrição foram realizadas conforme as instruções do fabricante New England Biolabs. Quando houve necessidade de defosforilação do vetor, foi utilizada a Shrimp Alkaline Phosphatase (SAP) (Promega) seguindo-se as instruções do fabricante. A fosforilação de fragmentos foi realizada com uma polinucleotídeo quinase proveniente do *kit* Transformer Site-Directed Mutagenesis (Clontech), de acordo com as instruções do *kit*.

Fragmentos de DNA eluídos de géis de agarose foram purificados com o *kit* Wizard SV Gel and PCR Clean-Up System (Promega) seguindo as orientações especificadas pelo fabricante. A união de fragmentos de DNA fita

simples para originar sequências dupla fita para clonagem foi realizada pela desnaturação e anelamento dos fragmentos: primeiramente, aquecendo-se o DNA fita simples numa solução 1x do tampão de reação da enzima Shrimp Alkaline Phosphatase (SAP) (Promega) a 98 °C por 5 minutos. Em seguida, deixando a solução voltar lentamente à temperatura ambiente.

No processo de construção de vetores por ligação, produtos de PCR ou fragmentos de DNA provenientes de digestão foram unidos ao vetor numa proporção de 3:1 quando provenientes de digestão e 10:1 quando provenientes de PCR. Para a ligação utilizou-se T4 DNA ligase (USB) de acordo com as instruções do fabricante.

Outro método de construção de vetores utilizado foi por recombinação *in vitro*. Este método usou os *kits* In-Fusion HD Cloning (Clontech) ou NEBuilder HiFi DNA Assembler Master Mix (New England Biolabs). Os fragmentos foram amplificados com sequências homólogas ao vetor. A reação foi realizada usando a mistura de reagentes fornecida pelos *kits* seguindo-se suas instruções.

A montagem de vetores *in vivo* em *K. phaffii*, descrita por Yu e colaboradores (2012), foi utilizada neste trabalho por meio de fragmentos com extremidades homólogas ao vetor replicativo na transformação da levedura por eletroporação.

4.4. Sequenciamento de plasmídeos

O sequenciamento para conferência da sequência dos plasmídeos construídos foi enviado para a empresa Myleus Biotecnologia (Belo Horizonte), que utiliza o método de Sanger para amplificação das sequências e a plataforma ABI 3730 (Thermo Scientific) para eletroforese.

4.5. Análises de bioinformática

O alinhamento simples de sequências utilizou a ferramenta *online* nucleotide Basic Local Alignment Search Tool - BLASTn (blast.ncbi.nlm.nih.gov/Blast.cgi). Já as sequências obtidas por sequenciamento dos vetores foram comparadas com as sequências esperadas

utilizando a ferramenta *online* Multiple Sequence Comparison by Log-Expectation (MUSCLE) situada no endereço www.ebi.ac.uk/Tools/msa/muscle. A busca por repetições invertidas utilizou a ferramenta EMBOSS einverted (www.bioinformatics.nl/cgi-bin/emboss/einverted).

Simulações de montagens de plasmídeos e análise de sequências utilizaram o *software* SnapGene Viewer.

4.6. Preparação de células bacterianas competentes

As células de *E. coli* competentes foram preparadas de duas maneiras diferentes, para transformação por choque térmico ou por eletroporação. A linhagem foi semeada em meio LB ágar a partir de um estoque armazenado a -80 °C e incubadas por 16 h a 37 °C. A seguir, um dos seguintes protocolos foi realizado:

4.6.1. Preparo de células bacterianas quimicamente competentes

Este protocolo foi adaptado de Sambrook e colaboradores (2001). Uma colônia isolada da bactéria foi inoculada em 10 mL de meio SOB em um frasco Erlenmeyer de 125 mL e incubada a 37 °C por 16 h sob agitação (200 rpm). Um a dois mL deste pré-inóculo foram adicionados a 100 mL de meio SOB e incubados por tempo suficiente para uma OD₆₀₀ de 0,3. A cultura foi então resfriada por 15 min em banho de água-gelo e as células foram submetidas a centrifugação a 3000 x *g* por 5 min a 4 °C. O precipitado foi ressuscitado em 32 mL de tampão de transformação I e incubado em gelo por 15 min. As células foram então coletadas por centrifugação e ressuscitadas em 4 mL de tampão de transformação II. Ao final do protocolo, as células foram separadas em alíquotas de 100 µL e congeladas em gelo seco. O armazenamento foi feito a -80 °C.

No procedimento de transformação foi utilizada uma alíquota de células competentes descongeladas em gelo. O DNA foi adicionado às células competentes e estas foram incubadas por 30 min no gelo. A seguir as células foram submetidas a um choque térmico de 42 °C por 90 segundos e novamente incubadas no gelo por 2 min. As células receberam 1 mL de meio

LB, foram recuperadas por 1 h a 37 °C e semeadas em meio LB ágar com o antibiótico adequado.

4.6.2. Preparo de células bacterianas para eletroporação

Este protocolo foi utilizado de acordo com as instruções do *kit* Transformer Site-Directed Mutagenesis (Clontech). Uma colônia isolada foi inoculada em 5 mL de meio LB. As células foram incubadas por 16 h a 37 °C sob agitação (200 rpm). Um mL deste pré-inóculo foi adicionado a 100 mL de meio LB num frasco Erlenmeyer de 500 mL. A cultura foi incubada a 37 °C sob agitação de 200 rpm até uma OD₆₀₀ de 0,5. O frasco foi então resfriado em gelo por 20 min e as células foram coletadas por centrifugação a 1200 x g por 5 min a 4 °C. Após a remoção de todo o sobrenadante, as células foram ressuspensas em 10 mL de 10% glicerol gelado e estéril. Em seguida, foi feita uma incubação em gelo por 20 min e as células foram novamente coletadas por centrifugação nas mesmas condições. O precipitado foi ressuspensado em 100 µL de 10% glicerol gelado e as células foram separadas em alíquotas de 40 µL, congeladas em gelo seco e armazenadas a -80 °C.

Na transformação por eletroporação foram usados os seguintes parâmetros numa cubeta de 0,1 cm: 1,8 kV, 10 µF e 600 Ω. Após o choque, as células receberam 1 mL de meio LB, foram recuperadas por 1 h a 37 °C e foram semeadas em meio LB com o antibiótico adequado.

4.7. Extração de plasmídeos de *E. coli* por lise alcalina

A extração dos plasmídeos de *E. coli* foi realizada pelo método de lise alcalina descrito por Sambrook e colaboradores (2001), com modificações. Uma colônia isolada da placa de transformantes foi inoculada em 5 mL de meio LB com o antibiótico adequado e este inóculo foi inculado a 37 °C por 16 h sob agitação. As células de 3 mL de cultura foram coletadas por centrifugação a 10000 x g por 2 minutos e o precipitado foi ressuspensado em 200 µL de solução TE. Foram adicionados 360 µL de solução II para provocar a lise das células e a amostra foi incubada a temperatura ambiente por 5 minutos. Posteriormente foram adicionados 300 µL de solução III gelada e a mistura foi

incubada por 5 minutos em gelo. A amostra foi coletada por centrifugação a 10000 x *g* por 5 minutos e o sobrenadante foi transferido para outro tubo, centrifugando novamente caso o sobrenadante não estivesse límpido. O sobrenadante foi adicionado a 750 µL de isopropanol, homogeneizado por inversão e submetido a centrifugação a 10000 x *g* por 5 minutos. O sobrenadante foi descartado e o precipitado ressuspendido em 200 µL de solução TE.

Foram adicionados 110 µL de acetato de amônio 7,5 M e a amostra foi homogeneizada em agitador do tipo vórtice. Em seguida, a amostra foi submetida a centrifugação a 10000 x *g* por 10 minutos, o sobrenadante foi transferido para um novo tubo e precipitado através da adição de 750 µL de etanol 100% gelado. Seguiu-se uma centrifugação a 10000 x *g* por 5 minutos, o descarte do sobrenadante e uma lavagem com 500 µL de etanol 70% gelado. Foi feita uma última centrifugação a 10000 x *g* por 2 minutos, o sobrenadante descartado e o precipitado foi então secado em um concentrador a vácuo e ressuspendido em 50 µL de água ultrapura contendo RNase A (20 µg/mL). A amostra foi incubada a 37 °C por 15 minutos. O DNA foi armazenado a -20 °C.

Quando necessário, também foram utilizados os *kits* de extração de plasmídeos Wizard SV Plus (Promega) e Qiagen Plasmid Maxi (Qiagen).

4.8. Mutagênese sítio-dirigida

A mutagênese sítio-dirigida do vetor contendo o gene *ADE3* foi realizada com o *kit* Transformer Site-directed Mutagenesis (Clontech) seguindo as instruções do fabricante.

4.9. Transformação de leveduras por eletroporação

A transformação de leveduras seguiu o protocolo indicado pelo manual *Pichia expression kit* (Thermo Fischer Scientific), com algumas modificações quando da utilização de plasmídeos integrativos: Uma colônia fresca de levedura foi inoculada em 10 mL de meio YPD em um frasco Erlenmeyer de 250 mL e incubada a 30 °C sob agitação de 200 rpm por 24 h. 100 a 1000 µL desta cultura foram utilizados para inocular 100 mL de meio YPD em frasco

Erlenmeyer de um litro, o qual foi incubado a 30 °C sob agitação (200 rpm) até atingir uma OD₆₀₀ entre 1,3 e 1,5. As células foram então mantidas em gelo e coletadas por centrifugação a 3000 x g por 5 minutos a 4 °C. Foram feitas duas lavagens (com 100 e 50 mL) de água destilada estéril gelada e uma com 4 mL de sorbitol 1 M gelado. Ao final, o precipitado foi ressuscitado em 100 µL de sorbitol 1 M gelado e 80 µL de células foram usados para a eletroporação com 5-10 µg de DNA. A eletroporação utilizou uma cubeta de 2 mm e parâmetros de 1500 V, 25 µF e 200 Ω. Um mililitro de 1 M sorbitol gelado e estéril foi imediatamente adicionado à cubeta e, no caso de plasmídeo com marca de seleção de resistência a antibiótico, as células foram recuperadas por 2 h a 30 °C.

No caso da transformação da levedura com plasmídeos replicativos, o protocolo de transformação foi adaptado a partir de Wu e Letchworth (2004). Um pré-inóculo de 10 mL foi inoculado e incubado a 30 °C sob agitação de 200 rpm por 16 h e o volume adequado (de 100 a 1000 µL) foi utilizado para inocular 100 mL de meio YPD, incubado sob as mesmas condições até uma OD₆₀₀ entre 1 a 2. Os 100 mL de cultura foram submetidos a centrifugação a 2000 x g por 5 minutos a 4 °C e as células foram ressuscitadas em 8 mL de solução de pré-tratamento. Após incubação à temperatura ambiente por 30 minutos, foi feita uma nova coleta de células por centrifugação. As células foram então ressuscitadas em 1,5 mL de sorbitol 1 M e transferidas para um tubo de 1,5 mL em gelo. Em seguida, as células foram submetidas a centrifugação a 2000 x g por 5 minutos e ressuscitadas em 1,5 mL de sorbitol 1 M três vezes. Finalmente, o precipitado de células foi ressuscitado em 100 µL de sorbitol 1 M e utilizado para a eletroporação. Um volume de 80 µL de células competentes foi misturado com 1-10 µg de DNA e transferido para uma cubeta de eletroporação de 2 mm. A eletroporação seguiu os parâmetros: 1500 V, 200 Ω, 25 µF. Imediatamente foi adicionado 1 mL de 1M sorbitol gelado à cubeta e a mistura foi transferida para um tubo de 1,5 mL que foi incubado a 30 °C por 2-3 horas para recuperação. As células foram então plaqueadas em meio seletivo.

4.10. Análise de estabilidade dos plasmídeos em meio líquido

Os plasmídeos contendo cada um dos centrômeros de *K. phaffii* foram testados quanto à estabilidade no crescimento da levedura em meio YPD sem pressão seletiva. A metodologia utilizou como base um protocolo realizado em *S. cerevisiae* (PANZERI; PHILIPPSEN; 1982). Uma colônia isolada de cada construção foi inoculada em 20 mL de meio YPD líquido. O crescimento foi realizado a 30 °C sob agitação de 200 rpm por 24 h. A partir deste pré-inóculo, foi iniciado o crescimento para análise de estabilidade sob as mesmas condições, em 20 mL de YPD começando com uma OD₆₀₀ de 0,1. A cada 24 h, as células foram inoculadas em 20 mL de meio YPD fresco a uma OD₆₀₀ de 0,1 e incubadas novamente. Este procedimento se repetiu até um total de 144 horas de cultivo.

Ao final de 96 e 144 horas de cultivo, as células foram diluídas numa taxa de 1:10⁶ em água destilada estéril e 100 microlitros foram semeados em placas de meio YPD sólido com e sem antibiótico. As placas foram incubadas por 72 h a 30 °C.

4.11. qPCR

As linhagens provenientes de LA3 transformadas com os plasmídeos centroméricos, assim como aquela transformada com o vetor replicativo pPICH-ADE3, foram cultivadas em meio YPD contendo zeocina, enquanto a linhagem controle LA3 foi cultivada em meio YPD não seletivo. Todas foram cultivadas a 28 °C sob agitação de 200 rpm até atingirem uma OD₆₀₀ de 1. Dez mililitros das culturas foram coletados por centrifugação a 2000 x *g* por 5 minutos. As células foram ressuspensas em 1 mL de SDS 0,25% e incubadas a 98 °C por 8 minutos de acordo com o protocolo descrito por Nakamura e colaboradores. Ao final, a mistura foi submetida a centrifugação e o sobrenadante foi diluído 10 vezes em água ultrapura para ser usado nas reações de qPCR.

As reações de qPCR foram realizadas de acordo com as instruções do reagente iTaq Universal SYBR Green Supermix (Bio-Rad). Foram usados *primers* qZEO-F e qZEO-R para a quantificação dos plasmídeos e qHIS-F e

qHIS-R para o controle interno de cópia única. Os experimentos utilizaram um termociclador Rotor-Gene Q (Qiagen). A análise utilizou o método de quantificação absoluta e as curvas padrão utilizadas abrangeram a faixa de 1×10^4 a 1×10^8 cópias do gene de interesse. Os vetores pPIC9 (Invitrogen) e pPICH foram linearizados e utilizados como padrões para as curvas padrão dos genes *HIS4* e *Sh ble*, respectivamente.

As curvas padrão de cada gene foram utilizadas para calcular os valores correspondentes, e o número de cópias dos plasmídeos foi calculado dividindo os valores obtidos para o gene *Sh ble* pelos valores obtidos para o gene *HIS4* (JOSHI et al., 2008).

Os experimentos foram realizados em triplicata e os resultados analisados com o *software* GraphPad Prism 5. O resultado correspondente a cada um dos vetores centroméricos foi comparado ao resultado obtido com o vetor pPICH-ADE3 através de um teste t ($p < 0,05$).

4.12. Eletroforese em gel de poliacrilamida não desnaturante

Os géis de poliacrilamida foram preparados a partir das soluções estoque descritas previamente para uma concentração final de 12% de acrilamida, de acordo com Sambrook e colaboradores (2001). Para cada gel separador, foram utilizados: 2 mL da solução 30% acrilamida; 1,25 mL de 1,5 M Tris (pH 8,8); 50 μ L de 10% APS; 3 μ L de TEMED e 1,7 mL de água destilada. Na camada de gel concentrador, foram usados 0,42 mL de 30% solução acrilamida; 0,34 mL de 0,5 M Tris (pH 6,8); 25 μ L de 10% APS; 3 μ L de TEMED e 1,75 mL de água destilada.

As linhagens de levedura foram crescidas em 10 mL de meio MD-Ac durante 24 h a 30 °C sob agitação de 200 rpm. Dois mililitros da cultura foram concentrados por centrifugação a 13.000 x g durante 3 min e as células foram pesadas (peso úmido). Cada 100 mg de células receberam 450 μ L de solução de lise. Posteriormente, foi adicionada a quantidade necessária de esferas de vidro de 0,5 mm de diâmetro (400 mg para cada 100 mg de peso úmido de células). A mistura foi submetida a agitação vigorosa por 30 segundos, seguida de incubação no gelo por 30 segundos. Este procedimento foi repetido 30 vezes. A parte líquida foi transferida para um tubo de 1,5 mL e submetida a

centrifugação a 10.000 x *g* por 30 min a 4 °C. O sobrenadante desta centrifugação foi transferido para um novo tubo e armazenado a 4 °C até o uso.

A quantidade adequada de sobrenadante, que variou de 10 a 18 µL, foi aplicada no gel de poliacrilamida e a corrida foi realizada a 90 V por 2 h. Foi utilizado o sistema de eletroforese mini-PROTEAN Tetra Cell da Bio-Rad. Após o término da corrida, a fluorescência da proteína EGFP foi detectada no equipamento Amersham Imager 600.

4.13. Eletroforese de Campo Pulsado (PFGE)

O procedimento de eletroforese de campo pulsado utilizou um protocolo adaptado de Ohi e colaboradores (1998). Amostras foram preparadas a partir de células de levedura cultivadas em 10 mL de meio YPD. Após 24 horas de cultivo a 28 °C sob agitação de 200 rpm, as células foram coletadas por centrifugação a 1200 x *g* por 5 minutos e o meio de cultura foi removido. As células foram lavadas em 1 mL de 0,05 M EDTA e novamente submetidas a centrifugação sob as mesmas condições. As células foram ressuspensas em 300 µL de uma solução contendo 0,05 M EDTA (pH 8,0) e 0,05 M DTT e incubadas a 30 °C por 20 minutos. Em seguida, foi realizada uma nova centrifugação e o precipitado foi ressuspensado em 490 µL de 0,05 M EDTA. Foram adicionados 10 µL de uma solução de liticase 20 mg/mL e 500 microlitros de agarose solubilizada em água a 1% (Agarose for Pulsed-Field Electrophoresis: Sample Preparation, Sigma-Aldrich). A mistura foi colocada nos moldes para amostras do sistema (CHEF Genomic DNA Plug kit, Bio-Rad) até solidificar a 4 °C. Os blocos de agarose foram incubados em solução de esferoplastos a 37 °C por 16 horas. Finalmente, foram lavados com 0,05 M EDTA e incubados em solução de protease a 50 °C por três dias. Para armazenamento a 4 °C, os blocos de agarose foram lavados com 0,05 M EDTA. No caso de corridas que utilizaram um marcador comercial, 5 µL deste foram misturados a 5 µL de agarose para preparo de amostras 1% solubilizada em água e aplicados diretamente no poço correspondente no gel.

A corrida de eletroforese de campo pulsado utilizou o equipamento CHEF-DR III Pulsed Field Electrophoresis System (Bio-Rad) em duas diferentes condições, ambas de acordo com o manual do equipamento. A

corrida mais curta durou 6 h, tempo de *switch* 10 segundos, 6 V/cm, 120°; a corrida mais longa durou 48 h, tempo de *switch* 700 segundos, 3 V/cm, 120°. A solução tampão utilizada para a corrida curta foi TBE 0,5x e para a corrida longa foi TAE 1x. Após a corrida, o gel foi incubado em solução do respectivo tampão de corrida com 1 µg/mL de brometo de etídio por 15 minutos. Em seguida, o gel foi lavado em água destilada por 10 minutos e exposto à luz ultravioleta para visualização.

4.14. Extração de DNA total de leveduras

O DNA genômico de *K. phaffii* foi obtido através de um protocolo de extração de DNA com solventes orgânicos adaptado de Sambrook e colaboradores (2001). Foram utilizados 40 mL de meio YPD com ou sem antibiótico em frascos de 250 mL a 30 °C sob agitação de 200 rpm durante 24 a 72 h. As células foram coletadas por centrifugação a 3.000 x *g* por 5 min e ressuspensas em 3 mL de tampão SE. Foram adicionados 100 µL de uma solução de liticase 20 mg/mL e a mistura foi incubada a 37 °C por 30 min a 1 h. As células foram coletadas por centrifugação a 3.000 x *g* por 5 min e ressuspensas em 3 mL de TE20. Foram adicionados 500 µL de SDS 10% e a mistura foi incubada a 65 °C por 30 min.

Após a incubação, foram adicionados 1,5 mL de acetato de potássio 5 M (pH 8,9) e o material foi incubado 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 novo tubo, onde recebeu 1 volume de uma mistura de fenol:clorofórmio (1:1). Após agitação suave, o material foi submetido a centrifugação a 3.000 x *g* por 10 min. A fase aquosa foi transferida para um novo tubo e a este material foram adicionados 2,5 volumes de etanol 100%. Após 5 min de incubação a temperatura ambiente a mistura foi coletada por centrifugação 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 ressuspensado em 100 µL de TE contendo RNase A (0,1 mg/mL).

4.15. Southern Blot

Aproximadamente 10 µg de DNA genômico foram digeridos com Sall a 37 °C por 16 h e o DNA digerido, assim como alíquotas do mesmo DNA intacto, foram aplicados em um gel de agarose 1%. O gel foi tratado com solução de depuração por 15 min, seguindo-se de duas lavagens com solução de desnaturação por 15 min e, finalmente, duas lavagens com a solução de neutralização por 30 min. Em seguida, o gel foi transferido para a membrana de nitrocelulose como descrito no protocolo de transferência seca de Sambrook e colaboradores (2001).

A marcação das sondas, 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 385 pb da região codante do gene *Sh ble* obtido por PCR utilizando os *primers* 5zeo e 3zeo; e um fragmento de 570 pb obtido a partir de PCR do gene *ADE3* com os *primers* a-s-F e a-s-R. A temperatura de hibridização foi de 55 °C. A detecção por quimioluminescência foi feita com uma exposição de 15 minutos no equipamento Amersham Imager 600.

5. Resultados e discussão

A fim de facilitar a compreensão dos resultados, estes foram divididos em três capítulos. O capítulo 1 trata da construção de linhagens de *K. phaffii* com duplas auxotrofias. O capítulo 2 trata da construção de vetores centroméricos e da análise do comportamento destes quando comparados a um vetor replicativo. Já o capítulo 3 ilustra a montagem de um cromossomo artificial linear para *K. phaffii*.

5.1. Capítulo 1: Construção de linhagens auxotróficas de *K. phaffii*

A utilização de YAC linear na levedura demanda duas marcas de seleção diferentes, pois a transformação será feita com dois braços independentes além do inserto de interesse. A correta montagem *in vivo* do cromossomo artificial será identificada através dos efeitos dos dois genes envolvidos na seleção. Por esse motivo, e para contribuir com novas construções em *K. phaffii*, foram construídas linhagens com dupla auxotrofia.

Além das marcas de resistência aos antibióticos zeocina, higromicina e G418, há hoje para uso na levedura algumas marcas auxotróficas para aminoácidos e bases nitrogenadas (DROCOURT *et al.*, 1990; LINCEREGHINO *et al.*, 2008; CEREGHINO *et al.*, 2001; NETT *et al.*, 2005; YANG *et al.*, 2014). Dentre as marcas de seleção auxotróficas, os genes *ADE2* e *ADE3* envolvidos na síntese de purinas possuem características adicionais que os diferenciam dos demais genes de vias de síntese.

Como observado na figura 2, o gene *ADE2* codifica para a enzima responsável pela conversão de 5-aminoimidazol ribonucleotídeo (AIR) em carboxiaminoimidazol ribonucleotídeo (CAIR). Células de leveduras que possuem uma deleção ou mutação no gene *ADE2* apresentam acúmulo da molécula AIR, que tem coloração avermelhada. Portanto, estas linhagens apresentam colônias vermelhas (STOTZ; LINDER, 1990). Em *S. cerevisiae*, a deleção adicional de genes cujas enzimas atuem em etapas prévias à conversão de AIR em CAIR impede que a via atinja a síntese do composto vermelho. Dessa forma, linhagens com dupla deleção são brancas como as linhagens selvagens (UGOLINI; BRUSCHI, 1996). Exemplos de genes de enzimas localizadas nestas etapas anteriores da via são *ADE3* e *ADE5-8*. O gene *ADE3* codifica para a enzima tetrahydrofolato sintase (THF-sintase) que catalisa a síntese da molécula 10-formil-tetrahydrofolato (10-formil-THF) (figura 2). Esta molécula participa da via da adenina numa etapa anterior à etapa catalisada pela enzima codificada pelo gene *ADE2* e interrompe o fluxo metabólico antes do acúmulo do composto vermelho (BARLOWE, APPLING, 1990).

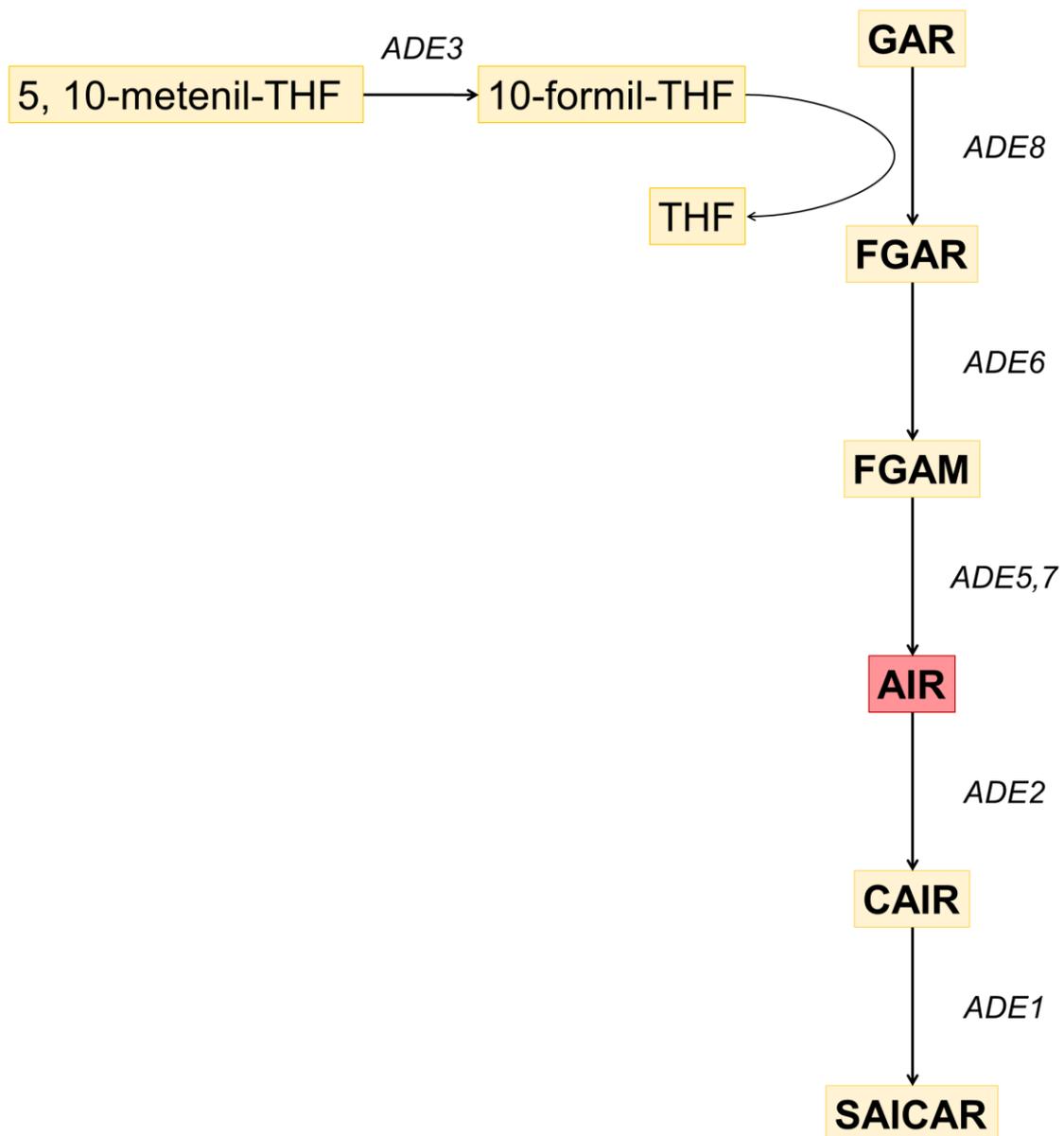


Figura 2. Via de biossíntese das purinas. Os intermediários retratados compreendem: GAR: glicinamida ribonucleotídeo; FGAR: *N*-formilaminoimidazol-4-carboxiamida ribonucleotídeo; FGAM: 5-fosforibosil-*N*-formilglicinamidina; AIR: 5-aminoimidazol ribonucleotídeo; CAIR: carboxiaminoimidazol ribonucleotídeo. SAICAR: *N*-succinil-4-carboxamida ribonucleotídeo. O composto AIR, responsável pela coloração vermelha, está indicado pela mesma cor.

A dupla deleção *ade2 ade3* nunca foi realizada em *K. phaffii*, embora já tenham sido realizadas mutações individuais nos genes *ADE2* e *ADE1* (DU; BATTLES; NETT, 2012). Também há um sistema de expressão disponível comercialmente pela empresa Thermo Fischer Scientific (*Pichia Pink*) baseado na auxotrofia causada pela mutação no gene *ADE2*.

A deleção do gene *ADE3* em *S. cerevisiae* causa um duplo efeito de auxotrofia. Além da auxotrofia para adenina, linhagens com este gene interrompido também são auxotróficas para o aminoácido histidina. Esta deleção interrompe a produção de 10-formil-THF que participa de duas etapas do metabolismo de purinas. A primeira foi mostrada na figura 2, juntamente com a glicinamida ribonucleotídeo, e a segunda faz parte de uma etapa relacionada ao metabolismo de histidina. Nesta, 10-formil-THF transfere seu grupamento formil para a molécula 5-aminoimidazol-4-carboxiamida (AICAR), um subproduto da via do aminoácido. Na falta de 10-formil-THF, AICAR passa a se acumular na célula. Esta molécula tem ação regulatória sobre a via de produção de histidina, embora o mecanismo ainda não tenha sido elucidado (RÉBORA; LALOO; DAIGNAN-FORNIER, 2005; TIBBETTS; APPLING, 2000).

5.1.1. Linhagem LA3

A linhagem *K. phaffii* LA2 (PIVA *et al.*, 2018), utilizada como ponto de partida para este trabalho, possui deleção do gene *AMD2* que foi realizada para facilitar o uso da marca de seleção *amdS* nesta levedura. O gene *amdS* de *Aspergillus nidulans* confere à levedura a habilidade de usar acetamida como fonte de nitrogênio. Além disso, esse gene pode ser utilizado em ensaios de contrasseleção em meio de cultura contendo fluoroacetamida: linhagens que possuírem a marca *amdS* morrem na presença deste composto (SOLIS-ESCALANTE *et al.*, 2013). A linhagem LA2 também possui uma deleção no gene *ADE2*, causando auxotrofia para adenina e coloração avermelhada para as colônias.

A partir desta linhagem, foram construídas outras com duplas deleções para teste com o cromossomo artificial. Assim, estariam disponíveis diversas marcas de seleção para os dois braços do YAC.

Primeiramente, o gene *ADE3* foi deletado na linhagem LA2. O cassete de deleção do gene *ADE3* foi construído por reações de PCR seguidas de uma reação de recombinação. Os pares de *primers* ADE3up-F/ADE3up-R e ADE3dw-F/ADE3dw-R amplificaram regiões de cerca de 600 pb flanqueando o gene *ADE3* no genoma de *K. phaffii* para induzir a recombinação no genoma por substituição. Já o par ADE3lox-F e ADE3lox-R permitiu a amplificação do

cassete *kan* de resistência a G418 flanqueado por sequências *loxP* a partir do vetor PGKL (REIS *et al.*, 2012). Os três fragmentos foram unidos ao vetor pBluescript II SK+ linearizado com *Sma*I por meio de recombinação *in vitro* com o *kit* In-Fusion.

O vetor resultante com o cassete de deleção foi utilizado como molde numa reação de PCR com os *primers* ADE3up-F e ADE3dw-R, amplificando todo o cassete. Este fragmento foi utilizado na transformação de *K. phaffii* LA2. Os transformantes foram selecionados baseando-se na resistência a G418 e pelo fenótipo de coloração branca. A confirmação da deleção do gene *ADE3* foi realizada por PCR de colônia com *primers* ADE3conF e ADE3conR, gerando um fragmento de 3000 pb (figura 3). A linhagem foi então chamada de LA3-Kan.

A linhagem LA3-Kan foi, em seguida, submetida a transformação com o vetor replicativo pYRCre2 (BETANCUR *et al.*, 2017) para a retirada do gene de resistência a antibiótico. Este vetor contém o gene que codifica a recombinase CreA sob controle do promotor *TEF1* de *S. cerevisiae*, capaz de retirar a marca de seleção que estava flanqueada por sequências *loxP* (Anexo 3).

Os transformantes foram selecionados em meio complexo com higromicina. Para confirmar a excisão da marca foi realizada uma PCR de colônia com *primers* ADE3conF e ADE3conR, gerando um fragmento de 1500 pb. Como controle, foi usada a linhagem LA2 (4500 pb)

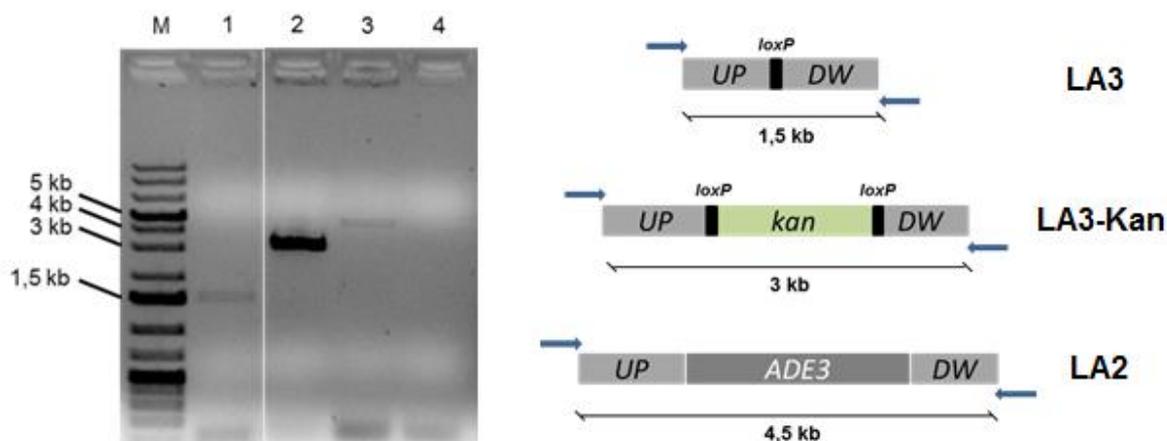


Figura 3. Deleção do gene *ADE3*. Eletroforese em gel de agarose 1% corado com brometo de etídeo exibindo PCR de colônia com *primers* ADE3conF e ADE3conR. Poço 1: linhagem LA3 com a marca de seleção reciclada. Poço 2: linhagem LA3-Kan

com a marca *kan* integrada no *locus ADE3*. Poço 3: Linhagem LA2 com gene *ADE3* intacto. Poço 4: Controle da reação de PCR. M: marcador Gene Ruler 1 kb Plus (Thermo Fischer Scientific).

Em *S. cerevisiae*, a deleção do gene *ADE3* interfere na regulação da via de síntese de histidina, causando um efeito de auxotrofia para este aminoácido. Para verificar se isso também ocorre em *K. phaffii*, a linhagem LA3 foi semeada em meio MD com adenina, adicionando ou não histidina. Foi observado crescimento no meio suplementado de histidina, enquanto no meio que continha apenas adenina a levedura não foi capaz de crescer (figura 4). Confirmou-se assim que o acúmulo de AICAR tem o mesmo efeito tóxico em *K. phaffii* que em *S. cerevisiae*, afetando múltiplas etapas da síntese de histidina (HÜRLIMANN *et al.*, 2011).

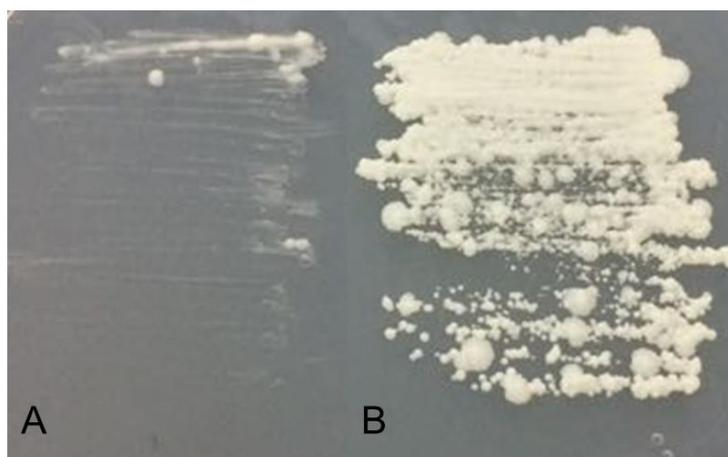


Figura 4. Análise fenotípica da linhagem LA3. A linhagem foi plaqueada em meio MD suplementado apenas com adenina (A) ou com adenina e histidina (B).

Linhagens com duplas deleções na via da adenina foram construídas em *S. cerevisiae* e utilizadas como ferramentas em estudos envolvendo construções heterólogas (THRASH-BINGHAM; FANGMAN, 1989). Uma linhagem *ade2 ade3*, branca, que recebe um vetor contendo o gene *ADE3* selvagem volta a ser vermelha enquanto mantiver este plasmídeo. Esta característica permite a verificação da estabilidade e do padrão de segregação desta construção (KOSHLAND; KENT; HARTWELL, 1985). Assim, a linhagem

LA3 foi utilizada na verificação da estabilidade dos plasmídeos construídos neste trabalho.

Para a utilização de um cromossomo artificial linear, fazem-se necessárias duas marcas de seleção. Com o efeito de auxotrofia para histidina, o gene *ADE3* passa a representar também uma marca de seleção auxotrófica na linhagem LA3.

5.1.2. Plasmídeo pPICH-Cre

O plasmídeo pPICH-Cre, contendo o gene que codifica a enzima CreA recombinase, foi construído em substituição ao plasmídeo pYRCre2. O vetor pYRCre2 possui a sequência de replicação autônoma de *S. cerevisiae* e, embora tenha sido utilizado na reciclagem de marcas anteriores, apresentou problemas em algumas transformações de *K. phaffii*. Em leveduras, tanto a eficiência de transformação quanto a estabilidade ao longo das divisões celulares podem ser prejudicadas pela sequência de replicação, visto que esta possui sequências espécie-específicas (CREGG *et al.*, 1985; SCHWARZHANS *et al.*, 2017). Por este motivo foi construído o vetor pPICH-Cre a partir do vetor pPICHOLI (MoBiTec). A origem de replicação deste vetor é a sequência ARS1 de *K. phaffii*.

Primeiramente, foi retirado um sítio de BamHI localizado dentro da sequência ARS1 e foi adicionado um sítio de AvrII no começo desta mesma sequência (figura 5A). Para isso, foi feita uma PCR com os *primers* pARS1 e pPIC-R, amplificando um fragmento de 2100 pb. Para adicionar sítios de restrição necessários às próximas etapas de clonagem (EcoRI, BamHI, NotI e PmeI), foi feita mais uma PCR utilizando este primeiro fragmento como molde. Esta PCR utilizou os *primers* pPIC-F e pPIC-R. A figura 5B mostra a região amplificada a partir do vetor comercial pPICHOLI com os sítios finais de restrição adicionados pelas reações de PCR. Estas duas reações amplificaram a região contendo a origem de replicação de *E. coli*, a marca de seleção *Sh ble* e a sequência ARS1 com os sítios de restrição necessários às etapas seguintes de construção do vetor pPICH-Cre e do cromossomo artificial (figura 5B).

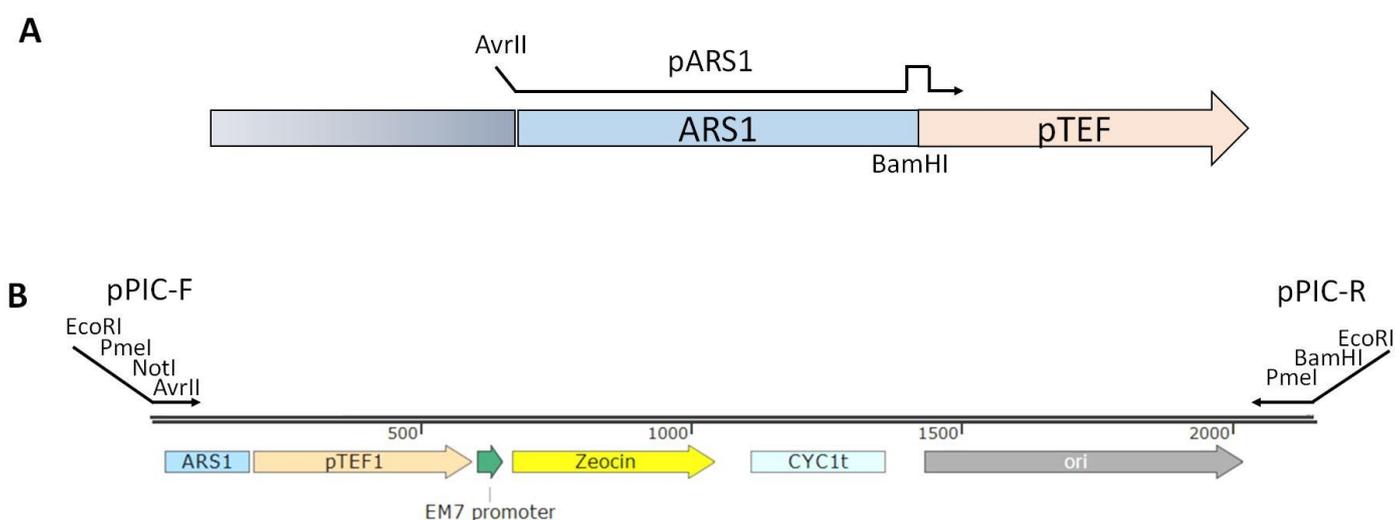


Figura 5. Representação da amplificação do vetor pPICHO1. A: Detalhe do sítio de anelamento do primer pARS1 na primeira reação de PCR. **B:** Detalhe do fragmento amplificado pela segunda reação de PCR, com *primers* pPIC-F e pPIC-R.

O fragmento obtido nesta PCR foi purificado e suas extremidades foram digeridas com EcoRI. O fragmento digerido foi religado para formar um vetor circular de aproximadamente 2100 pb. O vetor resultante foi chamado de pPICH.

O plasmídeo pPICH foi linearizado com a enzima de restrição NotI, defosforilado e ligado ao cassete de expressão da recombinase CreA, proveniente da digestão com NotI do vetor pYRCre2 (BETANCUR *et al.*, 2017). *E. coli* XL10-gold foi transformada com o sistema de ligação e foram feitas extrações de plasmídeos. O plasmídeo que apresentou um padrão de mobilidade eletroforética diferente dos demais quando intacto no gel foi digerido com a enzima EcoRV e deveria fornecer bandas de 2129, 1816 e 617 pb. O resultado é apresentado na figura 6.

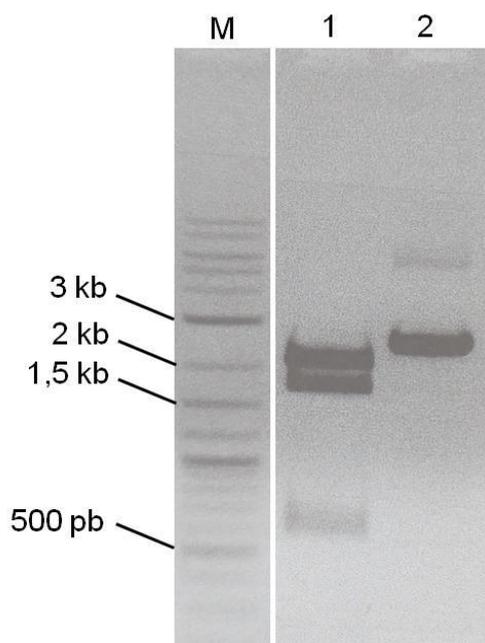


Figura 6. Confirmação da clonagem do cassete CreA no vetor pPICH. Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando digestão com EcoRV do vetor obtido. Poço 1: Plasmídeo digerido com EcoRV. Poço 2: Plasmídeo intacto. M: marcador 2-log DNA Ladder (New England Biolabs).

A digestão confirmou a presença do gene da CreA recombinase, portanto este foi chamado de pPICH-Cre e utilizado na reciclagem de marcas de seleção em *K. phaffii* (figura 7).

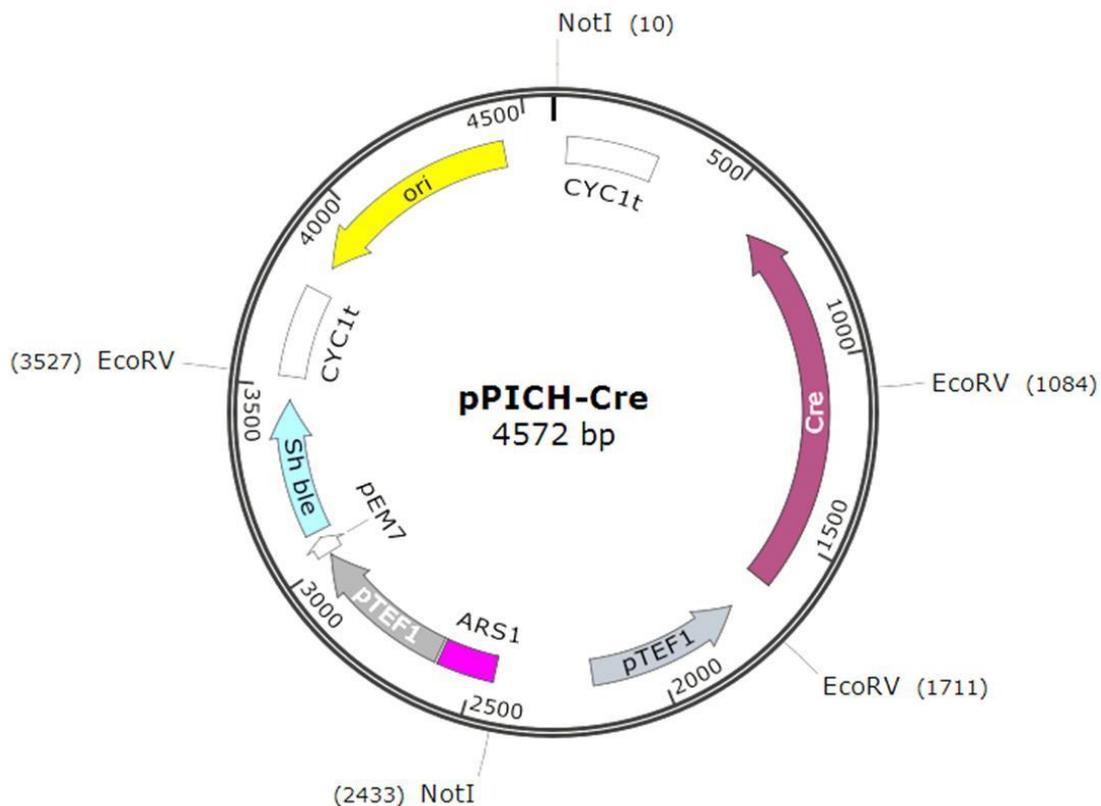


Figura 7. Mapa do vetor pPICH-Cre. Em rosa, sequência ARS1 de *K phaffii*. Em cinza, promotor *TEF1* de *S. cerevisiae*, em branco terminador *CYC1* de *S. cerevisiae*. Gene *Sh ble* de resistência a zeocina está em azul e CreA, em roxo. Origem de replicação bacteriana é indicada em amarelo.

5.1.3. Linhagem LAU

Para fornecer mais alternativas na construção do cromossomo artificial e em estudos futuros em *K. phaffii*, foram construídas linhagens adicionais com dupla auxotrofia. Uma linhagem com deleção no gene *URA5* foi construída utilizando a marca de seleção *amdS*. O gene *URA5* possui uma propriedade de contrasseleção com o ácido 5-fluorótico (5'-FOA). Este é convertido em um composto tóxico quando a enzima codificada pelo gene *URA5* está presente (NETT; GERNGROSS, 2003).

O vetor pURAMDS (PIVA *et al.*, 2015), ilustrado no anexo 5, foi utilizado para deleção por substituição do gene *URA5*. Este vetor contém sequências homólogas para recombinação de aproximadamente 1 kb com as regiões adjacentes ao gene *URA5* de *K. phaffii*. O vetor foi digerido com a enzima *SmaI*, utilizado na transformação de *K. phaffii* LA2 e a seleção foi feita em meio

MD-Ac. Dos clones obtidos em acetamida, 50 foram transferidos para uma placa de meio MD com 5-FOA e aqueles capazes de crescer neste meio indicavam a ausência do gene *URA5*. A confirmação final foi feita com uma PCR de colônia utilizando os *primers* URA5F e URA5R (figura 8 – poço 2), resultando na linhagem LAU-amdS.

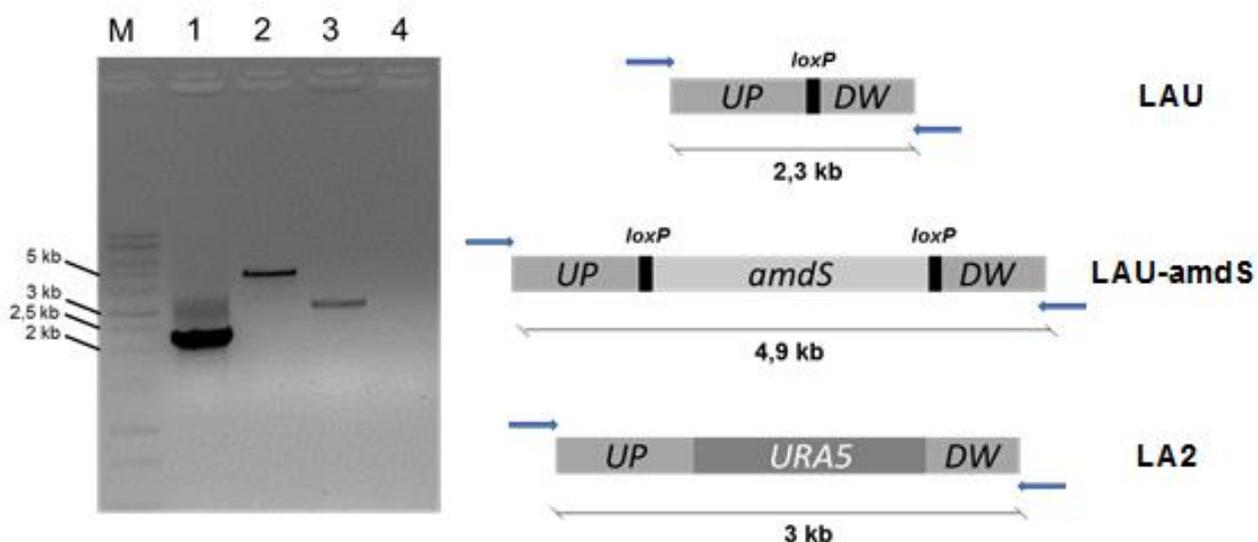


Figura 8. Deleção do gene *URA5*. Eletroforese em gel de agarose 1% corado com brometo de etídeo exibindo PCR de colônia com *primers* URA5F e URA5R. Poço 1: Linhagem LAU com marca de seleção reciclada. Poço 2: Linhagem LAU2-amdS com gene *URA5* substituído pela marca *amdS*. Poço 3: Linhagem LA2 com gene *URA5* nativo. Poço 4: Controle da reação de PCR. M: marcador 1 kb DNA Ladder (Promega).

A linhagem LAU-amdS foi transformada com o vetor pYRCre2 para a remoção da marca de seleção *amdS*. Entretanto, a reciclagem da marca não pôde ser realizada com este vetor, pois não foram obtidos clones na seleção da transformação em higromicina. Para contornar este problema, a levedura LAU-amdS foi então transformada com o vetor pPICH-Cre e selecionada em meio YPDS com zeocina. Clones isolados a partir desta placa foram semeados em meio MD com fluoroacetamida para buscar por leveduras que perderam a marca *amdS*. A partir de um clone que apresentou crescimento em fluoroacetamida, foi feita uma PCR de colônia para confirmação com os *primers* URA5F e URA5R (2,3 kb) e a linhagem final foi chamada de LAU (figura 8). Como controle, foi utilizada a linhagem parental LA2 (3 kb).

A linhagem LAU apresentou crescimento muito lento em meio YPD, recuperando-se em parte quando crescida em meio YPD suplementado com adenina e uracila. Linhagens de *P. pastoris* auxotróficas para adenina e/ou uracila apresentam crescimento mais lento, provavelmente devido a uma deficiência no transporte destes compostos para dentro da célula (CEREGHINO *et al.*, 2001). Além disso, duplas auxotrofias em leveduras podem afetar uma à outra, indicando que os efeitos das deleções são combinatórios e diversos no metabolismo da célula (VARSHNEY *et al.*, 2012).

Em vista deste resultado, foi construída mais uma linhagem com dupla auxotrofia: *ade2 leu2*. O cassete de deleção do gene *LEU2* foi obtido a partir do vetor pLEUΔKAN (anexo 4) (BETANCUR *et al.*, 2017) e utilizado para a transformação de *K. phaffii* LA2. As sequências utilizadas para recombinação por substituição foram de aproximadamente 650 pb. A seleção de colônias foi feita em meio YPD com G418. Dezoito colônias foram transferidas para placas de meio MD suplementado de adenina, com ou sem adição de leucina. Foi selecionado um clone para confirmação por PCR de colônia com *primers* PpLEU2-EXT1 e PpLEU2-EXT2. Esta linhagem foi chamada de LAL-Kan. A figura 9 mostra que a linhagem LAL-Kan forneceu um fragmento de 3300 pb, correspondente ao gene interrompido pelo cassete de seleção. A linhagem LA2 forneceu um fragmento de 2000 pb, correspondente ao gene nativo.

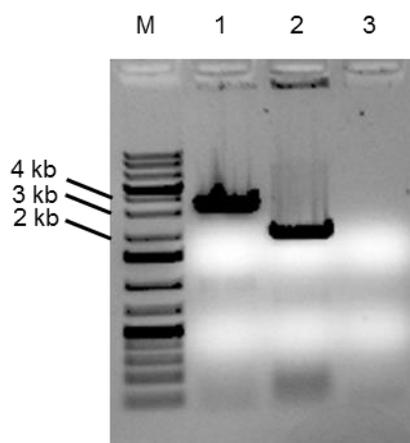


Figura 9. Deleção do gene *LEU2*. Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando PCR de colônia com *primers* PpLEU2-EXT1 e PpLEU2-EXT2. Poço 1: Linhagem LAL-Kan. Poço 2: Linhagem LA2. Poço 3: controle da reação de PCR. M: marcador Gene Ruler 1 kb Plus (Thermo Fischer Scientific).

A reciclagem da marca de resistência a G418 não teve sucesso e a linhagem LAL-Kan foi armazenada para testes futuros. As linhagens obtidas na primeira etapa deste trabalho forneceram alternativas para a construção do cromossomo artificial, visto que este necessita de duas marcas de seleção se for transformado de maneira linear com um inserto. Um resumo dos genótipos das linhagens construídas é exibido na tabela 3. A partir destes resultados, seguiu-se a construção do vetor contendo os componentes do cromossomo artificial.

Tabela 3. Linhagens construídas neste trabalho.

Linhagem	Genótipo
<i>K. phaffii</i> LA3	<i>amd2 ade2 ade3</i>
<i>K. phaffii</i> LAU	<i>amd2 ade2 ura5</i>
<i>K. phaffii</i> LAL-kan	<i>amd2 ade2 Δleu2::kan</i>

Parte dos resultados obtidos nesta etapa do trabalho foi publicada em um artigo na revista Applied Microbiology and Biotechnology sob o título “Acetamidase as a dominant recyclable marker for *Komagataella phaffii* strain engineering” (Anexo 8).

5.2. Capítulo 2: construção de vetores centroméricos

No genoma de organismos eucarióticos, os centrômeros são responsáveis pela segregação dos cromossomos na mitose e meiose e garantem a estabilidade do material genético ao longo das divisões celulares (NAGAKI *et al.*, 2012). Vetores de *S. cerevisiae* que contêm estas sequências apresentam uma taxa de segregação errônea durante as divisões celulares muito menor que a dos vetores replicativos comuns (HIETER *et al.*, 1985b). No caso de *K. phaffii*, um vetor centromérico contendo seu centrômero 2 foi recentemente construído e apresentou estabilidade aumentada quando comparado a um plasmídeo replicativo (NAKAMURA *et al.*, 2018). Além destas leveduras, plasmídeos centroméricos foram desenvolvidos para *Schizosaccharomyces pombe*, *Candida glabrata* e *Scheffersomyces stipitis* e em todos os casos apresentaram uma estabilidade mitótica elevada em condições não seletivas (CAO *et al.*, 2017; KITADA; YAMAGUCHI; ARISAWA, 1996).

A construção de vetores deste tipo para *K. phaffii* contorna a principal desvantagem dos plasmídeos replicativos, que é a instabilidade ao longo das divisões celulares em meios não seletivos (SCHWARZHANS *et al.*, 2017). Além disso, a clonagem dos centrômeros da levedura permitiria a caracterização funcional destas sequências antes da construção do cromossomo artificial. Portanto, vetores contendo cada um dos quatro centrômeros da levedura foram construídos para permitir a avaliação com relação à estabilidade do plasmídeo centromérico. A sequência centromérica que apresentasse os melhores resultados seria escolhida para a construção do YAC.

5.2.1. Construção de vetor replicativo contendo gene *ADE3*

A construção dos vetores replicativos partiu do vetor comercial pPICHOLI para aproveitar a sequência ARS1 de *K. phaffii* contida neste. Esta sequência de replicação autônoma é nativa de *K. phaffii*, encontra-se no

cromossomo 2 da levedura e tem sido utilizada em plasmídeos replicativos há mais de 30 anos (CREGG *et al.*, 1985).

O vetor pPICH foi linearizado com a enzima NotI e purificado para receber como inserto o gene *ADE3* amplificado a partir do genoma de *K. phaffii* X-33. Este gene foi amplificado juntamente com suas próprias regiões promotora e terminadora, representadas por 500 pb *upstream* e *downstream* da região codante. Para a PCR, foram utilizados os *primers* ADE3up-F e ADE3dw-R. O fragmento obtido foi digerido com NotI. Assim, o vetor de 2200 pb recebeu o gene em um inserto de pouco mais de 4000 pb. Após transformação de *E. coli* vários clones foram analisados por restrição com a enzima NotI e um deles apresentou o resultado esperado: fragmentos de 2200 pb e de 4000 pb, como pode ser visualizado nos poços 5 e 6 da figura 10. A digestão do vetor pode ter sido parcial, resultando nas bandas esperadas juntamente com uma banda remanescente do vetor intacto.

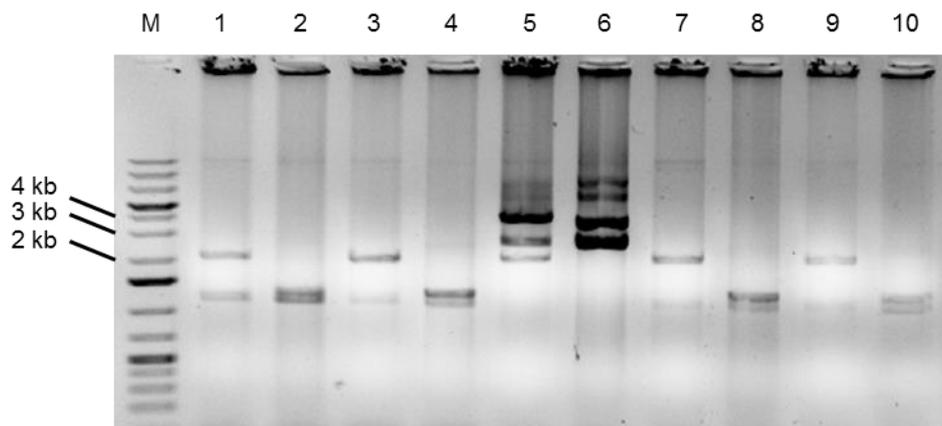


Figura 10. Confirmação da clonagem do cassete *ADE3* no vetor pPICH. Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão com NotI de plasmídeos extraídos de clones da transformação pPICH + *ADE3*. Poços 1, 3, 5, 7 e 9: plasmídeos digeridos com a enzima NotI. Poços 2, 4, 6, 8 e 10: plasmídeos intactos. M: marcador Gene Ruler 1 kb Plus (Thermo Fischer Scientific).

Adicionalmente, este plasmídeo teve a orientação da clonagem do gene *ADE3* confirmada por meio de digestões adicionais com as enzimas de restrição BamHI e EcoRI (figura 11).

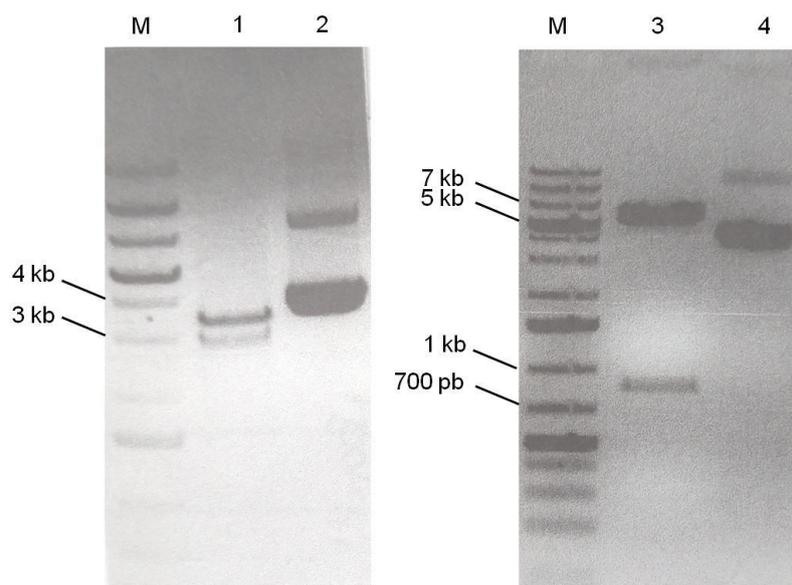


Figura 11. Determinação da orientação do gene *ADE3*. Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão do plasmídeo da transformação pPICH + *ADE3*. Poço 1: plasmídeo digerido com BamHI. Poço 2: plasmídeo intacto. Poço 3: plasmídeo digerido com EcoRI. Poço 4: plasmídeo intacto. M: marcador GeneRuler 1 kb Plus (Fermentas).

O resultado obtido por estas digestões (3,0 e 3,3 kb com BamHI; 850 pb e 5,5 kb com EcoRI) indica que a orientação do cassete *ADE3* estava como indicada na figura 12.

O próximo desafio foi trabalhar com os sítios de restrição disponíveis neste vetor. O gene nativo *ADE3* possui um sítio de restrição para a enzima BamHI em sua região codante, o que atrapalharia a clonagem da sequência centromérica no sítio de BamHI existente no vetor. Foi realizada então uma mutagênese sítio-dirigida utilizando os *primers* MUT1(Hpa) e MUT2(Bam). O *primer* MUT1(Hpa) foi utilizado como controle na reação de mutagênese e retirou um sítio de restrição da enzima HpaI localizado na posição 387 do vetor. O *primer* MUT2(Bam) retirou um sítio de BamHI localizado na posição 3395 do vetor, dentro da região codante do gene *ADE3*. A mudança não alterou o aminoácido correspondente ao códon. O vetor final foi chamado pPICH-ADE3 (figura 12).

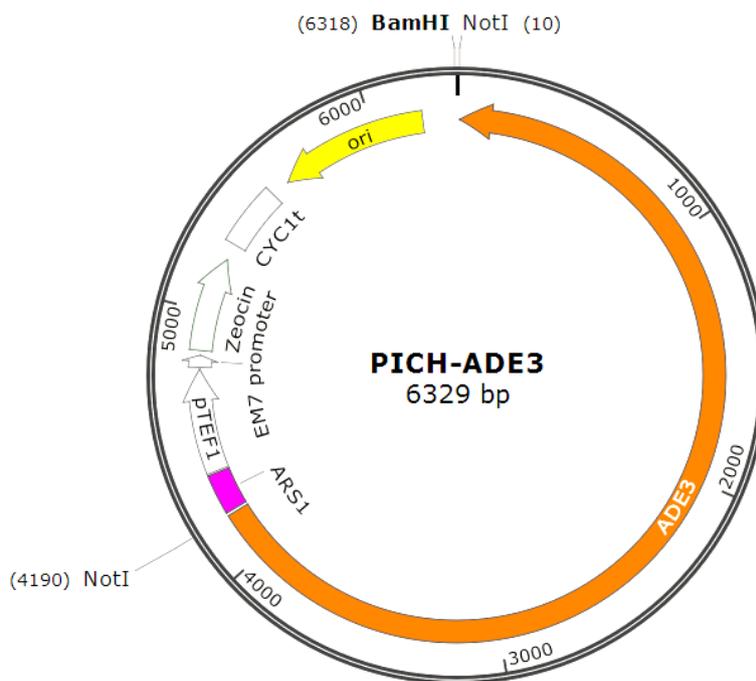


Figura 12. Mapa do vetor pPICH-ADE3. Em alaranjado, cassete *ADE3*. Em branco, cassete de expressão da marca de resistência a zeocina. Em rosa, sequência *ARS1*. Em amarelo, origem de replicação bacteriana.

5.2.2. Clonagem das sequências centroméricas

Visto que as sequências centroméricas de *K. phaffii* foram recentemente identificadas mas ainda não completamente caracterizadas, foi realizada a construção de plasmídeos com cada um dos centrômeros da levedura (NAKAMURA *et al.*, 2018). Dessa forma, seria possível analisar a funcionalidade dos centrômeros em um vetor circular, verificando se a estabilidade mitótica e o padrão de segregação do plasmídeo seriam compatíveis com o resultado esperado para um típico vetor centromérico.

O vetor pPICH-ADE3 foi utilizado para a clonagem das sequências centroméricas provenientes dos quatro cromossomos nativos de *K. phaffii*. O vetor foi linearizado com a enzima *Bam*HI (figura 12) para receber os fragmentos de centrômeros por recombinação *in vitro* com o *kit* In-Fusion. Os centrômeros foram amplificados por PCR e os *primers* desenhados a partir do genoma de *K. phaffii* GS115 (DE SCHUTTER *et al.*, 2009).

As regiões amplificadas de cada cromossomo apresentaram os seguintes tamanhos: 5488 pb para o centrômero 1, 6918 pb para o centrômero

2, 6693 pb para o centrômero 3 e o centrômero 4 com 6589 pb. Devido ao tamanho e à estrutura de repetição invertida destas sequências optamos por amplificar cada um dos centrômeros em duas partes de aproximadamente 3000 pb (figura 13). Os pares de *primers* utilizados foram CEN1-F/CEN1c-R e CEN1c-F/CEN1-R, e assim por diante para os outros centrômeros (CEN2: CEN2-F/CEN2c-R e CEN2c-F/CEN2-R; CEN3: CEN3-F/CEN3c-R e CEN3c-F/CEN3-R; CEN4: CEN4-F/CEN4c-R e CEN4c-F/CEN4-R).

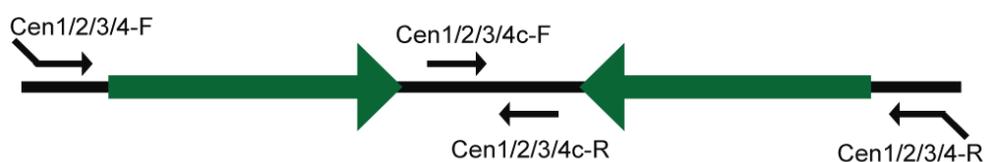


Figura 13. Anelamento dos *primers* para amplificação dos centrômeros de *K. phaffii*. Esquema representativo da região do genoma contendo os centrômeros. Setas verdes representam as sequências repetidas inversamente, setas pretas representam os *primers*.

Os vetores centroméricos foram montados por recombinação *in vitro* utilizando o *kit* In-Fusion (figura 14A). Os fragmentos possuíam regiões homólogas para recombinação entre si de ~80 pb. Para a recombinação dos fragmentos com o vetor pPICH-ADE3, foram utilizados 15 pb de homologia. Após a extração plasmidial e análise do perfil de restrição dos plasmídeos obtidos, observou-se que foram corretamente clonados os centrômeros provenientes dos cromossomos 1, 2 e 4 (figuras 14B, C e D, respectivamente). A digestão dos vetores obtidos nesta clonagem foi realizada com as enzimas NotI, PmeI e EcoRI e o padrão de bandas esperado após a digestão era de aproximadamente 8500 e 4300 pb com a enzima NotI, 6600, 3800 e 2500 pb com a enzima PmeI e 12000 e 800 pb com a enzima EcoRI. Como os centrômeros de *K. phaffii* apresentam tamanhos semelhantes, as digestões de todas as transformações deveriam fornecer fragmentos nestas faixas de tamanho. Os vetores obtidos nesta etapa foram chamados de pPICH-CEN1, pPICH-CEN2 e pPICH-CEN4. Não foi possível clonar o centrômero 3 por esta técnica, visto que a análise dos plasmídeos obtidos revelou que o vetor havia se religado ou apresentava eventos de recombinação.

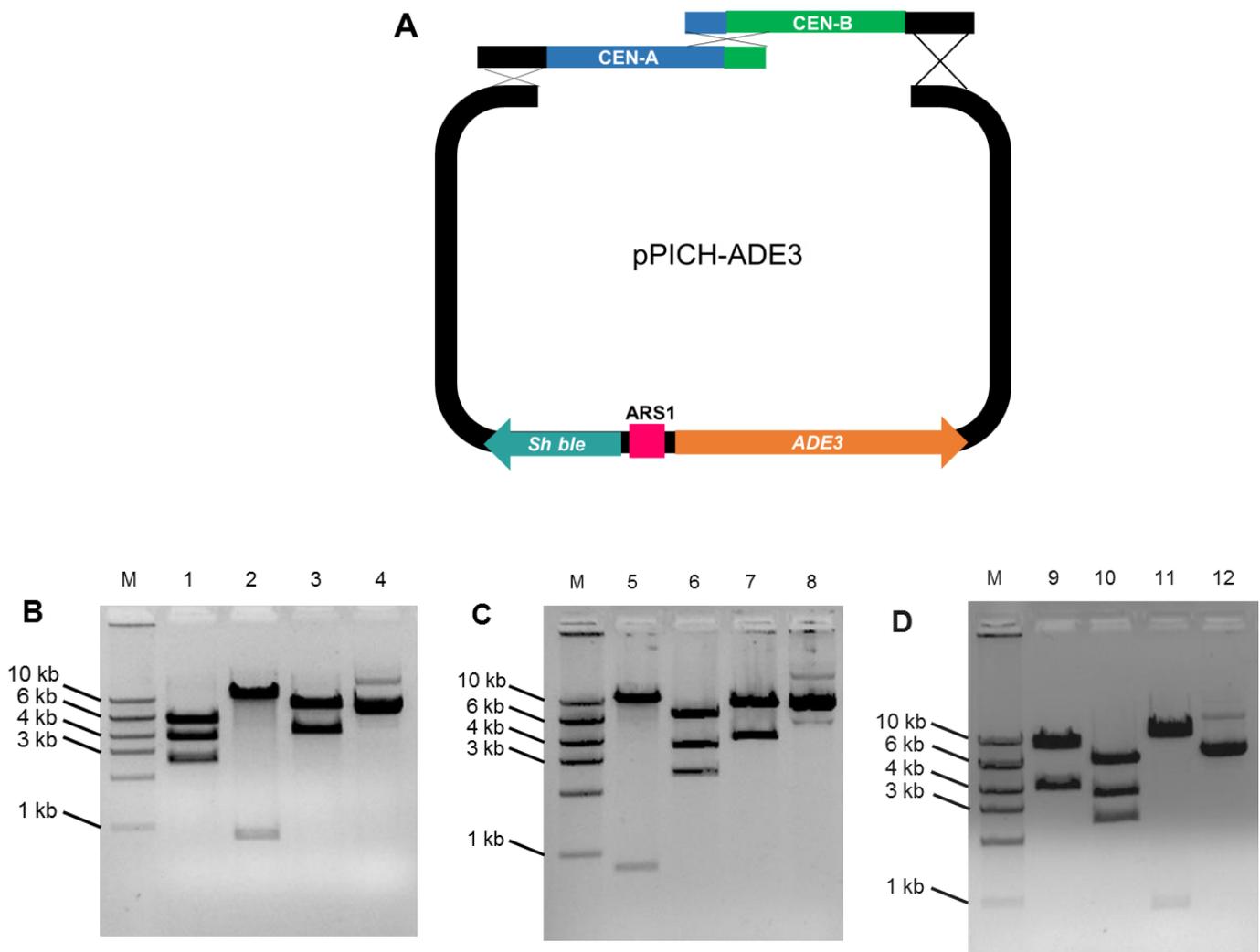


Figura 14. Confirmação da clonagem dos centrômeros. **A:** Estratégia de clonagem dos centrômeros em duas metades, CEN-A e CEN-B. **B:** Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão do plasmídeo pPICH-CEN1. Poços 1, 2 e 3: plasmídeo digerido com as enzimas PmeI, EcoRI e NotI. Poço 4: plasmídeo intacto. **C:** Digestão do plasmídeo pPICH-CEN2. Poços 5, 6 e 7: digestão do plasmídeo com as enzimas EcoRI, PmeI e NotI. Poço 8: plasmídeo intacto. **D:** Digestão do plasmídeo pPICH-CEN4. Poços 9, 10 e 11: digestão do plasmídeo com as enzimas NotI, PmeI e EcoRI. Poço 12: plasmídeo intacto. M: marcador High DNA Mass Ladder (Thermo Fischer Scientific).

A clonagem do centrômero 3 foi realizada por recombinação *in vivo* em *K. phaffii*. O centrômero 3 foi amplificado com os pares de *primers* CEN370-F/CEN3c-R e CEN3c-F/CEN370-R que adicionaram 70 pb de homologia entre os fragmentos do centrômero 3 e o vetor pPICH-ADE3. A levedura foi transformada com os fragmentos e mais o vetor pPICH-ADE3 linearizado com

BamHI. A transformação foi realizada na linhagem LA3 para permitir a verificação da presença do plasmídeo pela coloração avermelhada da colônia. Após seleção em zeocina, os clones obtidos foram isolados e armazenados para análises de estabilidade e de número de cópias do plasmídeo, assim como para a extração de DNA com o objetivo de verificar a integridade do plasmídeo circular, chamado de pPICH-CEN3.

Os vetores centroméricos pPICH-CEN1, pPICH-CEN2 e pPICH-CEN4 foram também transformados na linhagem *K. phaffii* LA3. Assim, a estabilidade mitótica dos quatro vetores poderia ser testada e comparada à do vetor replicativo pPICH-ADE3.

5.2.3. Análise de estabilidade dos plasmídeos centroméricos

A estabilidade mitótica dos 4 vetores centroméricos foi avaliada e comparada à do vetor replicativo pPICH-ADE3. O teste de estabilidade utilizou como parâmetro o fenótipo colorido das colônias. Colônias que mantêm o vetor são vermelhas, enquanto aquelas que o perderam voltam a ser brancas.

Dez clones de cada transformação foram selecionados e repicados em meio YPD com zeocina para a obtenção de colônias isoladas. Alguns clones de cada transformação perderam a capacidade de crescer em meio seletivo nesta passagem. Os clones restantes foram então semeados em meio YPD sem antibiótico e incubados a 30 °C por 3 dias. Esse crescimento foi realizado para verificar a presença do plasmídeo na ausência de pressão seletiva. Conforme mostrado na figura 15, 5 colônias do vetor pPICH-CEN1 foram isoladas e 3 mantiveram a coloração vermelha em meio não-seletivo. Três colônias da construção pPICH-CEN2 puderam ser isoladas, enquanto apenas uma manteve a coloração vermelha. Dez colônias da construção pPICH-CEN3 foram isoladas e todas mantiveram a coloração vermelha, enquanto que de 10 colônias da construção pPICH-CEN4, 8 estavam vermelhas. Em todas as placas, há uma colônia transformada com o vetor pPICH-ADE3 como controle.

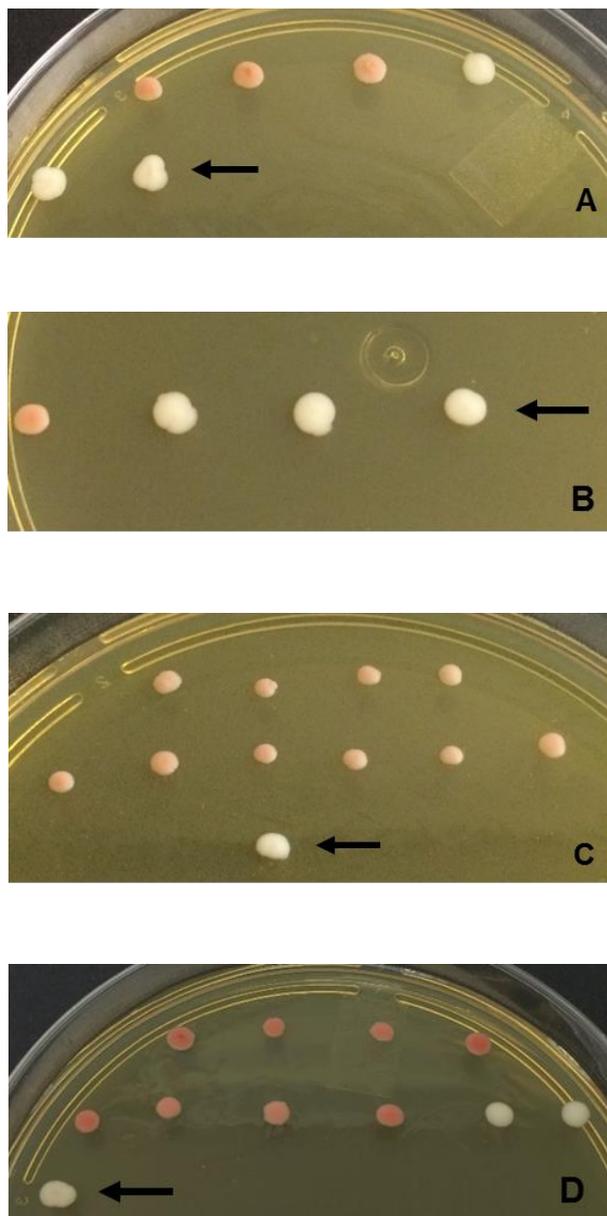


Figura 15. Fenótipo da linhagem LA3 transformada com vetores centroméricos. pPICH-CEN1 (A), pPICH-CEN2 (B), pPICH-CEN3 (C) e pPICH-CEN4 (D) em meio YPD sólido. Colônias indicadas por uma seta representam linhagem LA3 transformada com o vetor pPICH-ADE3.

Como se pode observar nas imagens, os vetores contendo as sequências centroméricas apresentaram maior estabilidade no crescimento em placa sem antibiótico que o vetor replicativo pPICH-ADE3. Colônias avermelhadas possuem o vetor contendo o gene *ADE3*, enquanto colônias brancas o perderam. Os vetores com melhores resultados foram pPICH-CEN3

e pPICH-CEN4, contendo os centrômeros dos cromossomos 3 e 4 de *K. phaffii*, respectivamente.

O próximo teste de estabilidade destes plasmídeos foi realizado em meio líquido, selecionando um clone que manteve a coloração avermelhada no teste em placa de cada transformação. Estes clones foram inoculados em meio YPD líquido e crescidos a 30 °C e sob agitação de 200 rpm, sendo diluídos e plaqueados em YPD com e sem antibiótico após 96 e 144 h de crescimento. As linhagens apresentaram crescimento semelhante e, para todos os cultivos, 96 horas de cultivo corresponderam a ~21 gerações, enquanto 144 horas a ~ 32 gerações. A figura 16 ilustra os resultados obtidos.

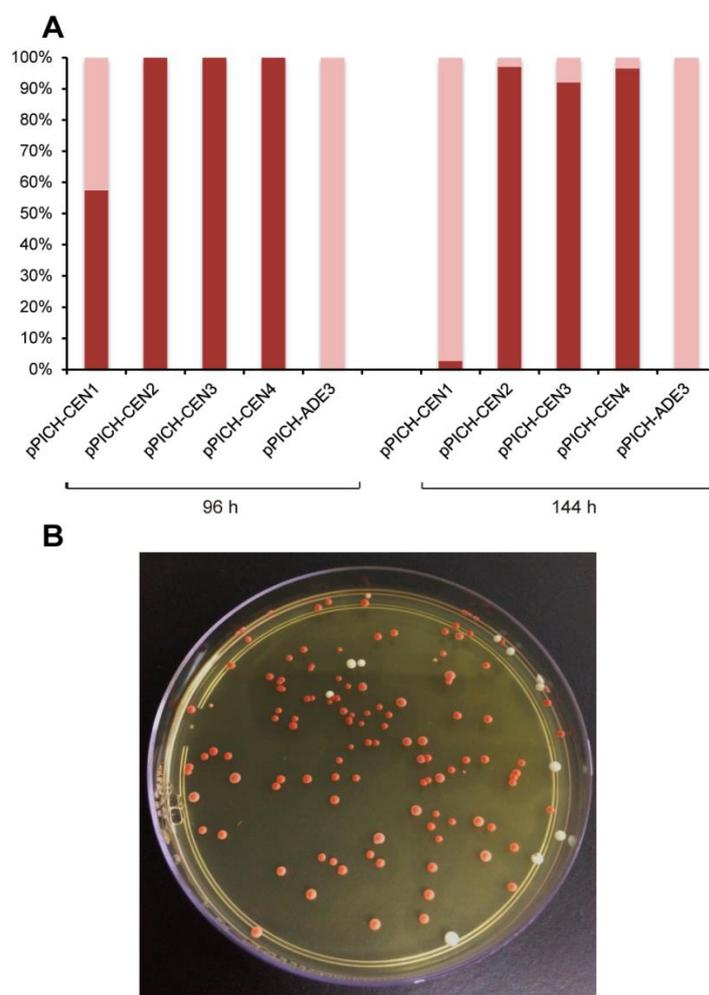


Figura 16. Teste de estabilidade mitótica. A: proporção entre colônias vermelhas (porção vermelha das barras) e colônias brancas (porção cor-de-rosa das barras) para cada plasmídeo transformado na linhagem LA3, após 96 e 144h de crescimento. **B:** exemplo de placa obtida após a diluição e o plaqueamento em YPD não seletivo.

Pode-se observar que após 96 h de cultivo a linhagem LA3 transformada com o vetor pPICH-ADE3, sem nenhuma sequência centromérica, já não possuía mais o plasmídeo. Suas colônias eram brancas (indicando ausência do gene *ADE3*). No caso dos vetores com sequências centroméricas, a estabilidade dos plasmídeos foi maior. O vetor pPICH-CEN1 forneceu 57% (27/47) de colônias vermelhas após 96 horas de cultivo, enquanto os vetores pPICH-CEN2, pPICH-CEN3 e pPICH-CEN4 forneceram 100% de colônias vermelhas. Após 144 horas, a linhagem transformada com os vetores pPICH-CEN2 e pPICH-CEN4 apresentou 97% (165/167 e 170/176, respectivamente) das colônias com plasmídeo. O vetor pPICH-CEN3 apresentou 92% (139/151) de colônias vermelhas, enquanto pPICH-CEN1 apenas 2,7% (3/110). Entretanto, estes dois últimos resultados apresentaram características inesperadas, visto que mesmo as placas com antibiótico apresentaram colônias brancas, o que indicaria ausência de plasmídeo.

Transformações utilizando a resistência a zeocina como marca de seleção podem exibir falsos positivos, ou seja, colônias resistentes ao antibiótico, mas que não expressam as características desejadas. Isso pode ocorrer pela degradação do antibiótico por longos períodos de incubação ou exposição à luz, assim como por mutações que criem resistência ou atrapalhem a expressão da proteína desejada (NORDÉN *et al.*, 2011). Neste trabalho, mutações podem ter prejudicado a expressão da enzima codificada pelo gene *ADE3* ou favorecido o desenvolvimento de linhagens resistentes a zeocina.

Os resultados obtidos em *K. phaffii* demonstram que os vetores pPICH-CEN1-4 possuem um fenótipo compatível com plasmídeos centroméricos, visto que garantem uma maior estabilidade quando comparados ao plasmídeo replicativo. O fato de o vetor pPICH-CEN1 ter apresentado a mais baixa estabilidade pode estar relacionado com a ausência de uma sequência ARS no centrômero 1, já que todas as outras sequências centroméricas amplificaram pelo menos parcialmente um elemento ARS endógeno da levedura (COUGHLAN *et al.*, 2016; LIACHKO *et al.*, 2014). O centrômero 2 foi amplificado com parte de um ARS no início de sua sequência (232 de 597 pb) e outra ao final (690 de 1189 pb). No fragmento do centrômero 3, foram amplificados 385 pb de um ARS que continha 816 pb ao total. Já no centrômero 4, um ARS de

637 pb foi completamente amplificada e uma outra sequência de 996 pb o foi parcialmente (840 pb) (figura 17).

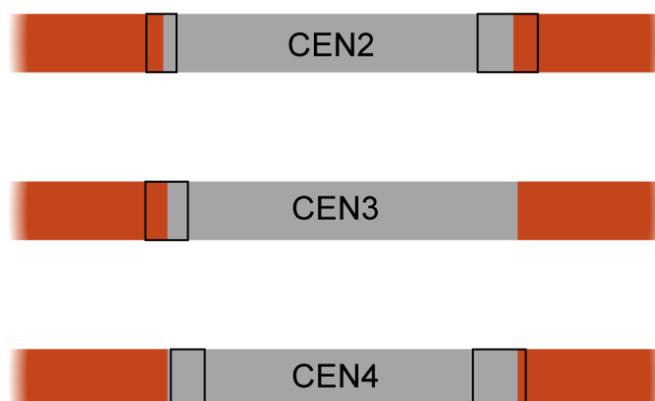


Figura 17. Representação esquemática das regiões centroméricas de *K. phaffii*. Em cinza, regiões centroméricas amplificadas e clonadas neste trabalho. Retângulos em preto indicam as regiões anotadas como ARS.

Sabe-se que a sequência ARS1, proveniente do cromossomo 2 de *K. phaffii* e contida nos vetores utilizados, contribui com uma menor eficiência na estabilidade mitótica de plasmídeos replicativos quando comparada a sequências mais recentemente descobertas (LIACHKO, DUNHAM, 2013). Portanto, a presença de sequências autônomas de replicação nos fragmentos centroméricos clonados nos vetores pPICH-Cen2-4 pode ter influenciado na maior estabilidade mitótica destes plasmídeos.

5.2.4. Análise do número de cópias dos plasmídeos centroméricos

Plasmídeos replicativos de leveduras são duplicados normalmente, porém, não são igualmente segregados na divisão celular, gerando algumas células com mais de uma cópia do vetor e outras que não o recebem. Quando submetidas a condições seletivas, as células sem plasmídeo não são capazes de crescer e as remanescentes constituem uma população que contém múltiplas cópias do plasmídeo (HIETER *et al.*, 1985). Dado que o centrômero é responsável pela ligação dos cromossomos duplicados ao fuso mitótico, o que garante a segregação destes na divisão celular, vetores centroméricos devem

fornecer um melhor padrão de segregação ao longo do crescimento da levedura e, conseqüentemente, um baixo número de cópias do plasmídeo (NAGAKI *et al.*, 2012; STEENSELS *et al.*, 2014).

A fim de verificar se este efeito é observado também no caso dos plasmídeos centroméricos de *K. phaffii*, foi realizado um ensaio de qPCR que utilizou um fragmento do gene de resistência a zeocina como indicador da presença dos plasmídeos, enquanto um fragmento do gene nativo *HIS4* foi usado como controle endógeno de uma cópia. Na figura 18, é possível verificar que o número de cópias dos vetores centroméricos foi mantido baixo (1 a 2) em todos os casos, enquanto o vetor pPICH-ADE3 esteve presente em ~ 25 cópias por célula. Vale ressaltar que as linhagens foram crescidas em meio seletivo, portanto células que perdessem o plasmídeo não seriam capazes de crescer.

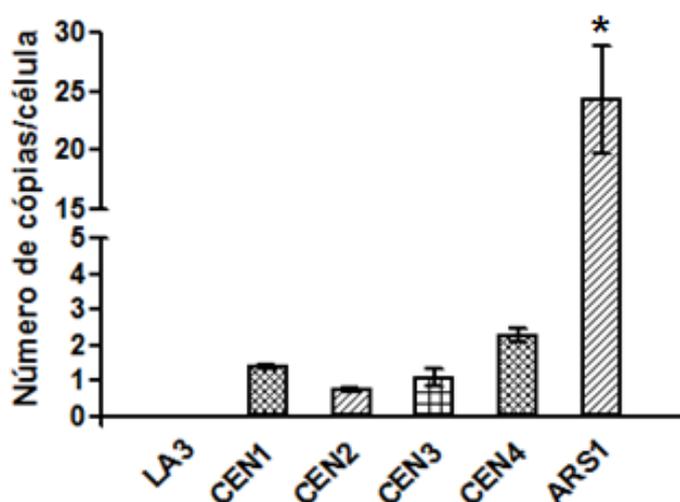


Figura 18. Determinação do número de cópias dos plasmídeos por célula. Linhagem LA3 pura e transformada com pPICH-CEN1, pPICH-CEN2, pPICH-CEN3 e pPICH-CEN4. ARS1: linhagem transformada com plasmídeo pPICH-ADE3. Barras de erro representam o desvio padrão da amostra (n=3).

A diferença entre o número de cópias dos vetores centroméricos e do vetor replicativo foi significativa ($p < 0,05$) e o resultado obtido foi condizente com o previsto para vetores centroméricos. Plasmídeos corretamente segregados mantêm-se mais estávelmente na célula e se dividem igualmente entre as células filhas, portanto permanecem presentes em um baixo número de cópias por célula. Caso o vetor seja erroneamente segregado durante a

divisão celular, este se acumula e cria células com múltiplas cópias e células sem o plasmídeo, que desaparecem sob condições seletivas.

Entretanto, este resultado não correspondeu ao obtido pelo trabalho que construiu um vetor contendo o centrômero 2 (NAKAMURA *et al.*, 2018). O resultado obtido pelo estudo mostra que o vetor replicativo está presente na célula em uma cópia, assim como o vetor centromérico e um vetor integrado ao genoma. Em *S. cerevisiae*, vetores centroméricos apresentaram diferenças em número de cópias com relação a um vetor replicativo quando marcas auxotróficas foram utilizadas nas construções. Quando a marca de seleção utilizada foi o gene *kan*, o número de cópias não variou entre vetores centroméricos e replicativos (KARIM; CURRAN; ALPER, 2013). Isto indica que outros fatores, além da origem de replicação, podem estar envolvidos no número de cópias de um plasmídeo circular em levedura e, portanto, deve-se considerar todos os aspectos de uma construção neste tipo de análise. Embora ambos utilizem a sequência autônoma de replicação ARS1 e a marca de seleção *Sh ble*, o vetor construído por Nakamura e colaboradores possuía um cassete de expressão do gene *GFP* e os vetores pPICH-CEN1-4 possuem o cassete *ADE3*.

O vetor pPICH-CEN4 foi um dos que apresentou melhor resultado quanto à estabilidade mitótica e ao número de cópias, além de apresentar um menor número de sítios de restrição em sua sequência. Portanto, o centrômero do cromossomo 4 foi escolhido para as próximas etapas de construção do cromossomo artificial.

5.2.5. Verificação da integridade dos vetores centroméricos

K. phaffii realiza recombinações não-homólogas com frequência e pode integrar plasmídeos replicativos em seu genoma, assim como formar plasmídeos a partir de sequências a princípio integradas (CREGG *et al.*, 1985; SCHWARZHANS *et al.*, 2016). Regiões de homologia com o genoma da levedura representam um risco de integração nos locais correspondentes (MADSEN; VISMANS; BRINCH-PEDERSEN, 2016). Por esta razão, a conformação circular dos vetores centroméricos pPICH-CEN1-4 foram verificadas antes de prosseguir com a construção do cromossomo artificial.

O DNA genômico dos clones transformantes com cada um dos vetores centroméricos foi extraído após cultivo em meio YPD com zeocina a 30°C sob agitação de 200 rpm por 72 horas. Este DNA foi transformado por eletroporação em *E. coli* XL10-gold e os plasmídeos recuperados através de extração por lise alcalina, demonstrando que a conformação dos vetores em *K. phaffii* era de fato circular.

Os vetores pPICH-CEN1, pPICH-CEN2 e pPICH-CEN4 foram recuperados através deste protocolo, enquanto o vetor pPICH-CEN3 não forneceu colônias após a transformação de *E. coli* e plaqueamento em meio seletivo. Este resultado pode indicar que *E. coli* não é capaz de replicar e manter o vetor pPICH-CEN3, visto que a montagem deste vetor não teve sucesso na bactéria. Entretanto, pode também indicar que o vetor pPICH-ADE3 linearizado integrou-se no genoma da levedura ao invés de participar da recombinação homóloga para a montagem do vetor pPICH-CEN3. Para verificar se este era o caso, foi realizado um experimento de Southern Blot com o DNA proveniente do clone pPICH-CEN3 digerido com a enzima NotI. De acordo com a figura 19, o vetor pPICH-CEN3 corretamente montado e digerido com esta enzima deveria fornecer bandas de 4180 pb e 8849 pb quando digerido com NotI.

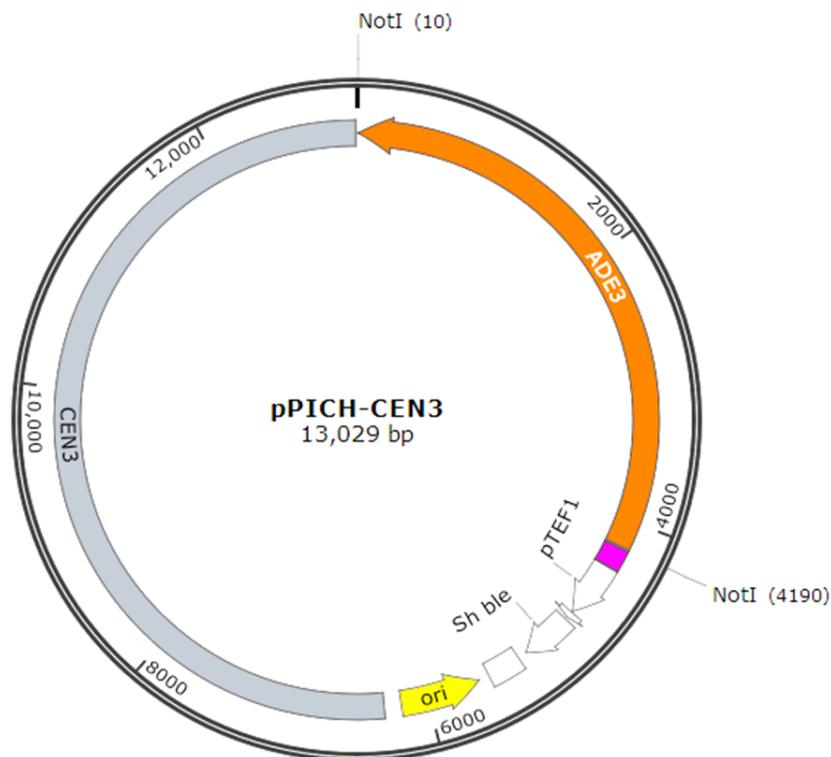


Figura 19. Mapa do vetor pPICH-CEN3. Em cinza está indicado o centrômero 3. Em alaranjado, o cassete *ADE3*. O cassete de resistência a zeocina é indicado em branco, a origem de replicação de *E. coli* em amarelo e ARS1 em rosa. Sítios de restrição da enzima NotI estão evidenciados.

No experimento de Southern Blot foi utilizada uma sonda que hibridiza ao gene *Sh ble* de resistência a zeocina, devendo revelar a banda de 8849 pb correspondente ao fragmento contendo o centrômero 3 da levedura. O resultado obtido (figura 20) revelou uma banda de ~4000 pb, indicando que o vetor não havia sido corretamente montado pela técnica de montagem *in vivo*. Este resultado, juntamente com o resultado negativo obtido nas transformações de *E. coli*, indica que o vetor pPICH-ADE3 provavelmente integrou-se no genoma da levedura e não recebeu os fragmentos do centrômero 3 por recombinação homóloga.

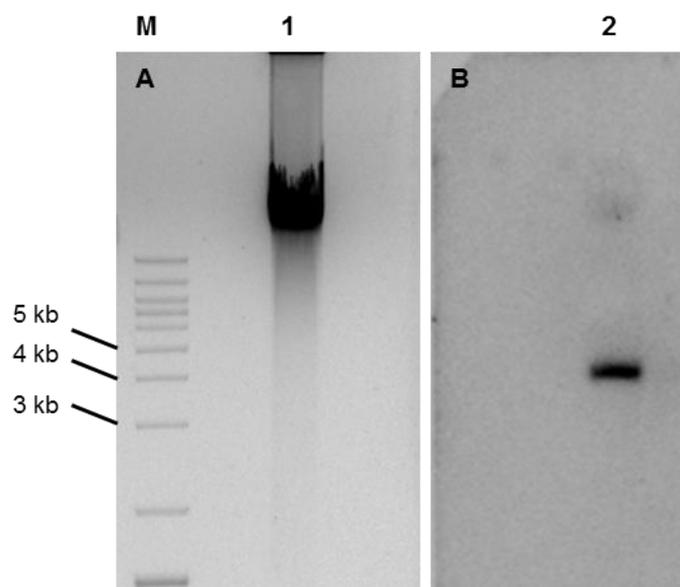


Figura 20. Southern Blot de DNA isolado da célula transformada com vetor pPICH-CEN3. A: eletroforese em gel de agarose 1% corado com brometo de etídeo. M: marcador 1 Kb Plus (Thermo Fischer Scientific). 1: DNA total digerido com NotI. **B:** Southern Blot hibridizado com a sonda *Sh ble*. 2: DNA total digerido com NotI e hibridizado com a sonda *Sh ble*.

Em vista deste resultado e considerando que o centrômero 4 foi escolhido para uso no cromossomo artificial, foram feitas reações de PCR do DNA total do clone transformado com o plasmídeo pPICH-CEN4. Estas reações foram uma confirmação adicional da conformação deste clone antes de iniciar a montagem do cromossomo artificial. As PCRs amplificaram fragmentos que se sobrepõem ao longo de todo o vetor utilizando os *primers* listados na tabela 4:

Tabela 4. Pares de *primers* e tamanho de fragmentos amplificados na verificação da conformação do vetor pPICH-CEN4.

Reação de PCR	Par de <i>primers</i>	Tamanho do fragmento (pb)
1	cBIB-R e CENc-F	3299
2	cBIB-F e CENc-R	3435
3	pPIC-R e 5zeo	1464
4	3zeo e mut(Bam)	1833
5	ADE3up-F e ADE3dw-R	4172 (plasmídeo) 1217 (genoma)

As reações de PCR com estes primers produziram fragmentos que percorrem todo o plasmídeo e, portanto, confirmam a conformação circular do vetor. Os resultados obtidos são ilustrados na figura 21. A reação número 5 produziu um fragmento proveniente do gene *ADE3* completo contido no plasmídeo e um segundo fragmento menor, proveniente da região do gene deletado no genoma.

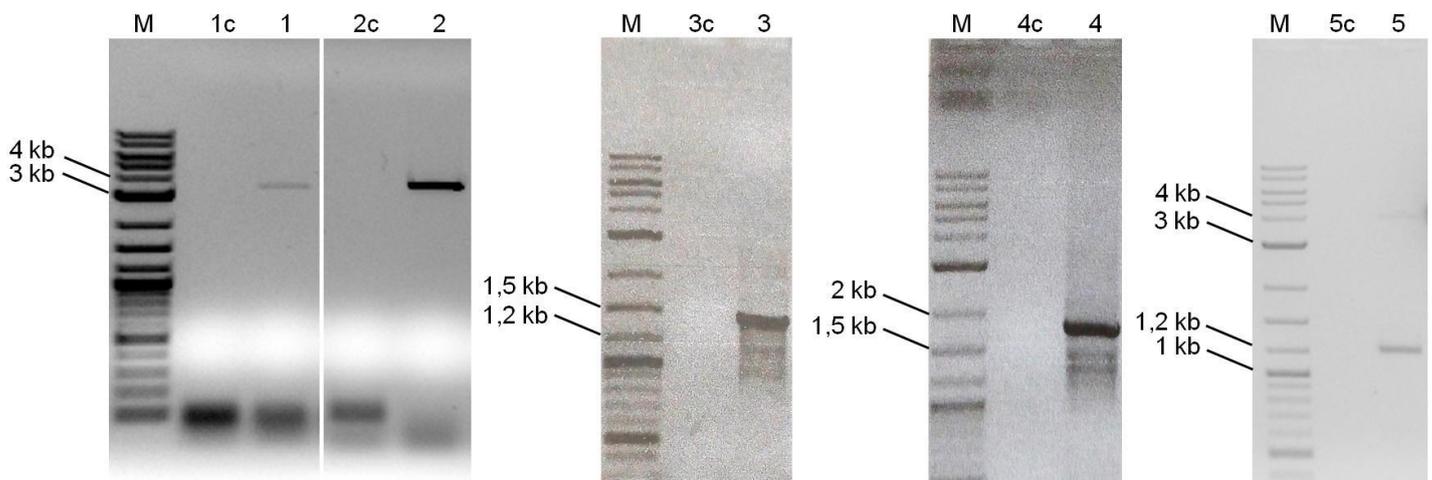


Figura 21. Confirmação da conformação do vetor pPICH-CEN4. Eletroforese em gel de agarose 1% corado com brometo de etídeo apresentando PCRs de confirmação do clone pPICH-CEN4. 1-5: reações 1 a 5 de acordo com tabela 4. 1c-5c: controles negativos das PCRs. M: marcador 2-log DNA Ladder (New England Biolabs).

Uma vez confirmada a conformação circular do vetor pPICH-CEN4, foi iniciada a construção do cromossomo artificial utilizando a sequência centromérica do cromossomo 4.

Parte dos resultados obtidos neste capítulo compôs um artigo submetido na revista *PLOS One*, sob o título “Functional characterization of *Komagataella phaffii* centromeres by a color-based plasmid stability assay”. O texto submetido está no anexo 9.

5.3. Capítulo 3: construção do cromossomo artificial

Os componentes básicos de um cromossomo artificial linear são um centrômero, duas marcas de seleção, uma sequência de replicação autônoma e dois telômeros em suas extremidades. Marcas de seleção e sequências de replicação autônoma já foram utilizadas em *K. phaffii* em diversas estratégias, sendo ferramentas bem caracterizadas (AHMAD *et al.*, 2014). Inúmeras novas ARS foram identificadas no genoma da levedura e constituem um conjunto de sequências disponível para uso em novos vetores (LIACHKO *et al.*, 2014).

Já centrômeros e telômeros foram recentemente identificados e inicialmente caracterizados, abrindo caminho para a construção de vetores contendo estas sequências. Os centrômeros dos quatro cromossomos de *K. phaffii* foram identificados e o centrômero 2 foi caracterizado em um plasmídeo circular (COUGHLAN *et al.*, 2016; NAKAMURA *et al.*, 2018). Repetições teloméricas foram identificadas nas extremidades de seus cromossomos através de estudos de genômica, porém ainda não foram utilizadas em um vetor linear (LOVE *et al.*, 2016).

Um vetor deste tipo concederia à levedura a capacidade de expressar, em menos tempo e com menos passos de clonagem, diversos genes heterólogos ao mesmo tempo. A quantidade de produtos sintetizados por engenharia metabólica em *K. phaffii* aumenta conforme a quantidade de ferramentas para sua manipulação se desenvolve (SCHWARZHANS *et al.*, 2017b). Além disso, um cromossomo artificial permitiria a continuidade do desenvolvimento de estratégias não integrativas na levedura, assim como do estudo dos níveis de expressão de proteínas (CAMATTARI *et al.*, 2016; SCHWARZHANS *et al.*, 2017).

À luz dos resultados obtidos nos capítulos 1 e 2, foi iniciada a construção de um cromossomo artificial contendo a sequência centromérica do cromossomo 4 de *K. phaffii*, sequências teloméricas, duas marcas de seleção e a sequência ARS1.

5.3.1. Amplificação e clonagem da sequência telomérica

De acordo com as sequências disponíveis no banco de dados GenBank para o genoma de *K. phaffii*, os telômeros são compostos por sequências repetitivas de cerca de 200 pb, variando em tamanho entre os cromossomos (DE SCHUTTER *et al.*, 2009; STURMBERGER *et al.*, 2016). As extremidades dos cromossomos desta levedura contêm uma repetição *in tandem* da sequência GCTGGAT. Devido à natureza repetitiva e à posição terminal desta sequência, a estratégia inicial de amplificação por PCR a partir do DNA genômico não teve sucesso para a obtenção do fragmento correspondente ao telômero e foi substituída pela estratégia descrita a seguir.

A segunda estratégia, elaborada para amplificação da sequência telomérica por PCR, utilizou como molde inicial um fragmento sintético representado pelo oligonucleotídeo TEL-PP (figura 22).

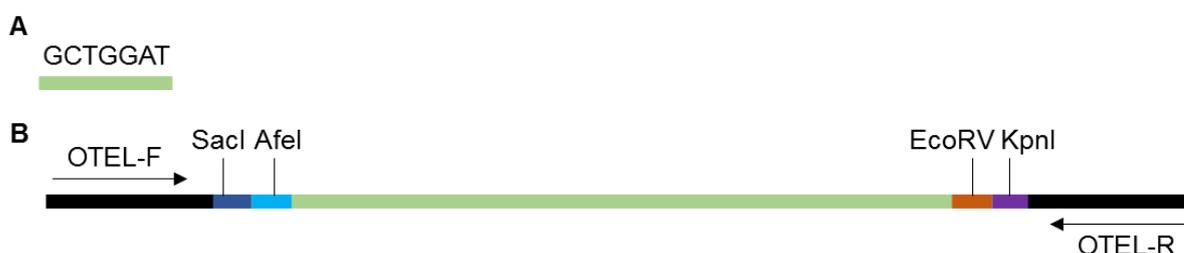


Figura 22. Representação do oligonucleotídeo TEL-PP. **A:** representação da região repetitiva do oligonucleotídeo. **B:** *primer* TEL-PP com a região repetitiva em verde, sítios de restrição e região de anelamento dos *primers* OTEL-F e OTEL-R (setas).

O fragmento TEL-PP foi sintetizado com 105 pb de repetições, ou seja, 15 módulos da sequência, e utilizado como molde em uma reação de PCR com os *primers* OTEL-F e OTEL-R. Este produto de PCR foi unido ao vetor pBluescript II SK+ linearizado com Sacl e KpnI por recombinação *in vitro* através do *kit* In-Fusion. As sequências OTEL-F e OTEL-R foram usadas como regiões de homologia para a recombinação.

Após a confirmação da clonagem do fragmento contendo as repetições teloméricas, o vetor foi submetido a duas digestões: a primeira, com EcoRV, para linearizá-lo a fim de receber novos insertos; a segunda, com EcoRV e

Afel, para liberar o fragmento de 105 pb para ser inserido no vetor previamente digerido com EcoRV. Assim, o vetor defosforilado recebe mais uma cópia do fragmento e dobra o tamanho da sequência repetitiva. Este par de enzimas de restrição foi escolhido por produzir extremidades abruptas que permitem a ligação de fragmentos em clonagens subsequentes sem alteração da sequência repetitiva. Após a ligação, ambos os sítios são perdidos, porém a repetição da sequência se mantém. A figura 23 mostra as sequências dos sítios de restrição e o resultado da ligação dos fragmentos digeridos.

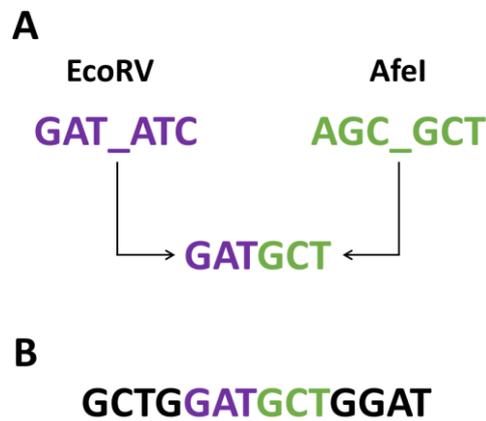


Figura 23. Sítios de restrição de EcoRV e Afel usados na ligação das repetições teloméricas. A: Sítios de EcoRV e Afel e resultado da ligação de fragmentos digeridos com estas enzimas. **B:** repetições teloméricas contendo a região da ligação dos fragmentos.

Assim, o vetor contendo 105 pb de região repetitiva foi preparado para receber mais fragmentos repetitivos em clonagens sequenciais. A clonagem sequencial foi repetida utilizando o vetor já com os 210 pb de repetições, obtendo um plasmídeo com 420 pb de sequência repetitiva e assim até obter um vetor com 840 pb de sequência telomérica (figura 24).

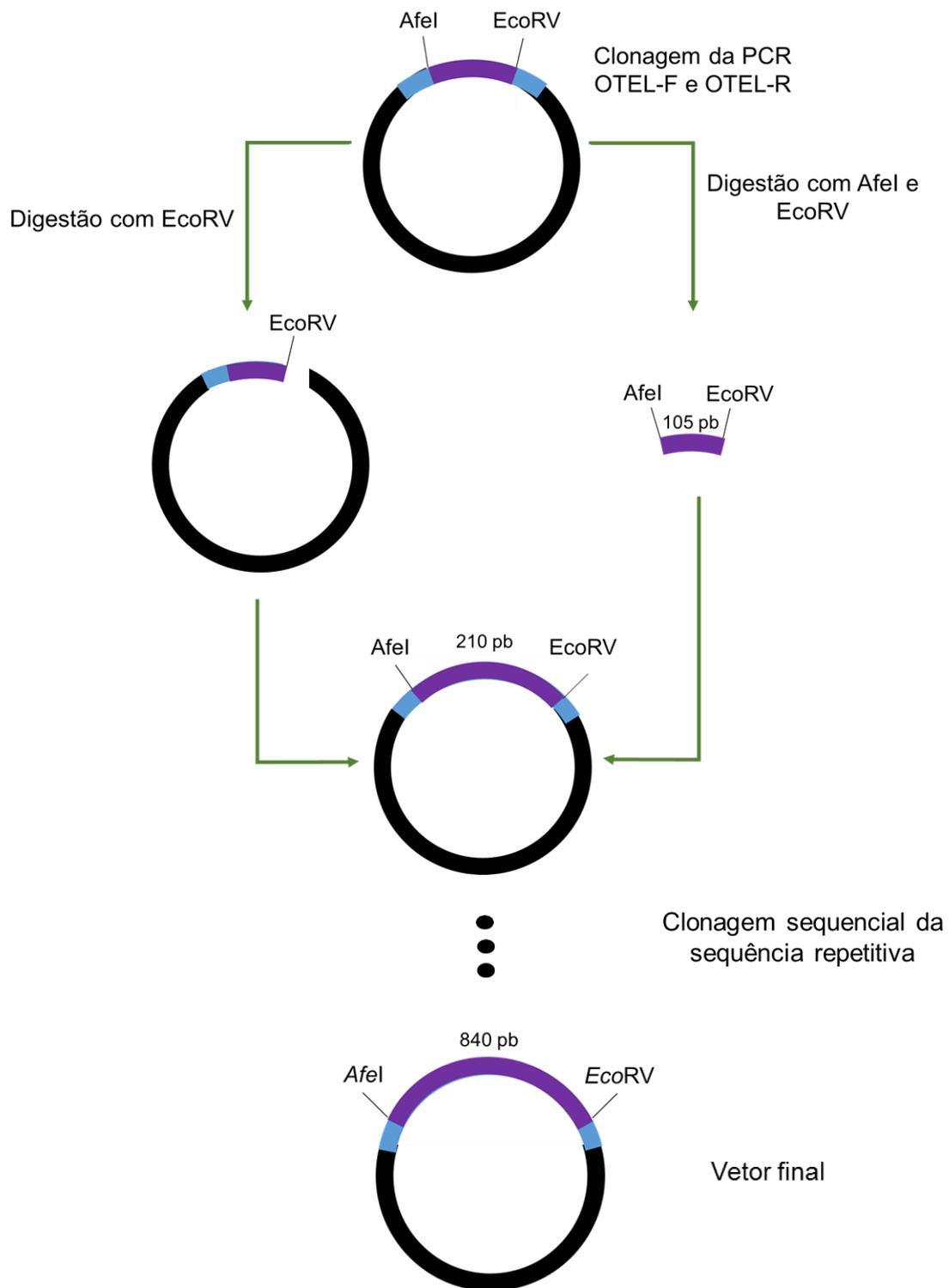


Figura 24. Estratégia da clonagem sequencial do fragmento contendo a repetição telomérica.

A clonagem foi confirmada por digestão com as enzimas de restrição Afel e EcoRV. A digestão correta libera um fragmento correspondente ao total

da sequência repetitiva, visto que os sítios intermediários se perdem caso a ligação tenha ocorrido corretamente. A confirmação foi realizada em cada etapa de clonagem sequencial (figura 25). O vetor pBlue-tel1 é o plasmídeo com a primeira clonagem (105 pb) da repetição telomérica, pBlue-tel2 o vetor com a clonagem da repetição duplicada (210 pb), pBlue-tel4 é o vetor com 4 vezes a repetição inicial (420 pb) e pBlue-tel8 é o vetor final (840 pb).

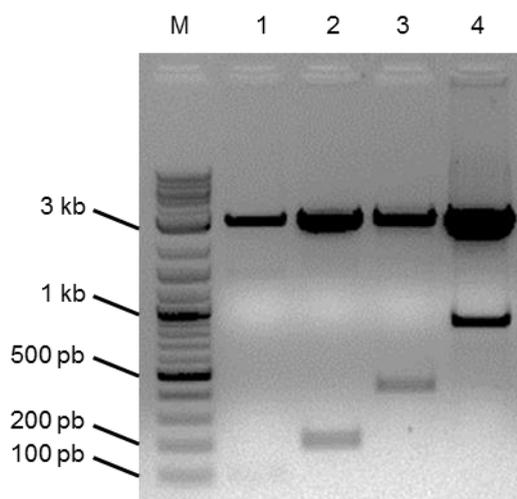


Figura 25. Confirmação da clonagem das repetições teloméricas. Eletroforese em gel de agarose 1% corado com brometo de etídeo. Digestão com Afel e EcoRV de plasmídeos oriundos do pBluescriptII SK+ com a sequência telomérica clonada sequencialmente. Poço 1: pBlue-tel1; poço 2: pBlue-tel2; poço 3: pBlue-tel4; poço 4: pBlue-tel8. M: marcador 2-Log DNA Ladder (New England Biolabs).

5.3.2. Análise da sequência telomérica clonada

O plasmídeo pBlue-tel8, com um inserto de 840 pares de bases de sequência repetitiva, foi enviado para sequenciamento utilizando os *primers* M13-F e M13-R. O resultado do sequenciamento foi comparado ao resultado teórico da montagem da sequência e este alinhamento é mostrado no anexo 1. Em vários pontos houve resultados discrepantes entre as duas reações de sequenciamento, o que pode ser explicado pela baixa qualidade dos picos observados nos primeiros nucleotídeos da reação de sequenciamento e pela natureza repetitiva da sequência.

Em dois pontos, o sequenciamento indicou a inserção de três pares de bases, porém as duas reações não indicaram os mesmos nucleotídeos.

Sequências repetitivas frequentemente apresentam problemas para amplificação e, considerando que o sequenciamento realizado utilizou o método de Sanger que se baseia na amplificação da amostra, provavelmente, estes resultados consistem em erros do sequenciamento (HOMMELSHEIM *et al.*, 2014).

5.3.3. Preparo do vetor pBluetel-8

A montagem final dos braços do cromossomo foi realizada utilizando duas extremidades teloméricas provenientes do vetor pBlue-tel8, o centrômero 4 proveniente do vetor pPICH-CEN4 e as marcas de seleção *Sh ble* e *ADE3*. Estas marcas de seleção foram escolhidas pela facilidade na clonagem e disponibilidade das sequências, assim como pela possibilidade de verificação da estabilidade do YAC ao final da transformação do vetor. Uma vez que a funcionalidade do cromossomo artificial estiver comprovada, devem ser utilizadas duas marcas de seleção auxotróficas nas linhagens previamente construídas.

O cromossomo artificial construído neste trabalho possui uma estrutura linear, logo precisa de dois telômeros em suas extremidades. Não foi possível realizar a clonagem das sequências teloméricas no vetor pPICH-CEN4, portanto todos os outros componentes do cromossomo artificial foram transferidos para o vetor pBlue-tel8f. Devido ao tamanho total do vetor e à dificuldade de clonagem das sequências, a montagem dos braços do YAC ocorreu em duas metades: o vetor pBlue-tel8 foi utilizado como base nos dois procedimentos de clonagem. Os dois vetores resultantes liberariam os braços lineares do cromossomo a partir de uma digestão com enzimas de restrição.

Primeiramente, foi necessário adicionar sítios de restrição no vetor pBlue-tel8. Um oligonucleotídeo foi sintetizado, hibridizado para formar uma sequência dupla fita e fosforilado. Este fragmento possuía um sítio de NdeI em seu interior e extremidades compatíveis com o vetor linearizado com SacI. Sua construção está ilustrada na figura 26.



Figura 26. Estratégia para construção do sítio de NdeI. Em azul, a sequência sintetizada e hibridizada para formar uma dupla fita. Em vermelho, extremidades do sítio de restrição da enzima SacI. O sítio de NdeI está indicado por um retângulo.

O vetor pBlue-tel8 foi linearizado com SacI e defosforilado para ser submetido a ligação com o fragmento construído e transformado em *E. coli*. Foram feitas 9 extrações de plasmídeos e o DNA foi digerido com NdeI, para verificar a existência do novo sítio na sequência (figura 27).

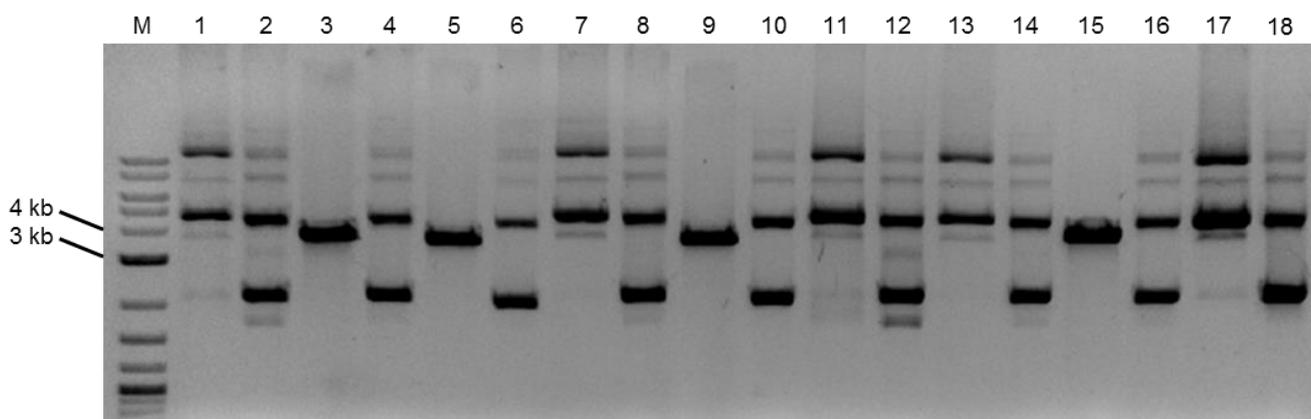


Figura 27. Confirmação da clonagem do fragmento contendo sequência NdeI. Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão de plasmídeos extraídos de clones da transformação pBlue-tel8+oligo NdeI. Poços 1, 3, 5, 7, 9, 11, 13, 15 e 17: plasmídeos digeridos com NdeI. Poços 2, 4, 6, 8, 10, 12, 14, 16 e 18: plasmídeos intactos. M: marcador 2-Log DNA Ladder (New England Biolabs).

As digestões dos poços 3, 5, 9 e 15 exibiram o padrão desejado, ou seja, o plasmídeo linearizado para aproximadamente 3800 pb. Um destes clones foi selecionado e armazenado para as próximas etapas de clonagem. O vetor resultante foi denominado pBlue-tel8N.

Em seguida foi elaborada uma estratégia para adicionar sítios das enzimas BamHI (para clonagem do centrômero 4), NotI (para clonagem do cassete de expressão *ADE3*) e mais um sítio de NdeI para permitir a liberação

dos braços lineares do YAC. Além destes sítios, foi também adicionado um sítio de *AvrII* para permitir a clonagem do promotor induzido *AOX1* que regularia a atividade do centrômero.

Esta estratégia consistiu na amplificação de uma região do vetor pBlue-tel-8N com *primers* pKpn e pPOLI. O *primer* pPOLI adicionou os sítios de *BamHI*, *AvrII*, *NotI* e *NdeI* ao fragmento amplificado (figura 28). Paralelamente, o vetor pBlue-tel-8 foi digerido com *NaeI* e *KpnI* para receber o novo fragmento por recombinação *in vitro* com o *kit* NEBuilder Assembly, então o vetor recebeu o mesmo fragmento, porém com os sítios de restrição adicionais.

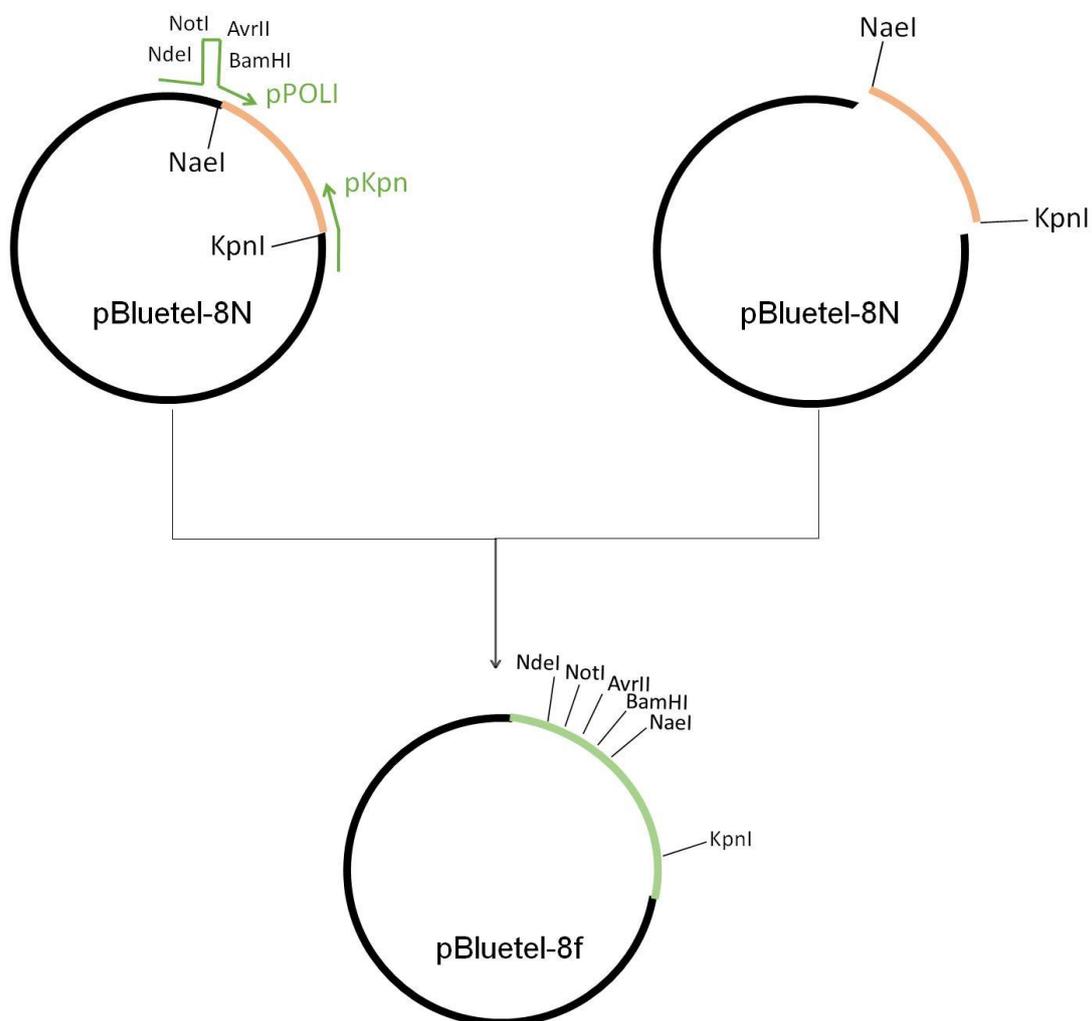


Figura 28. Estratégia de PCR e recombinação *in vitro* para adição de sítios de restrição ao vetor pBlue-tel8N. Em rosa, fragmento retirado do vetor pBlue-tel8N por digestão com *NaeI* e *KpnI*. Em verde, fragmento amplificado por PCR e clonado novamente com os novos sítios de restrição. O vetor final foi chamado de pBlue-tel8f.

A construção foi confirmada por digestão com a enzima de restrição NdeI (figura 29).

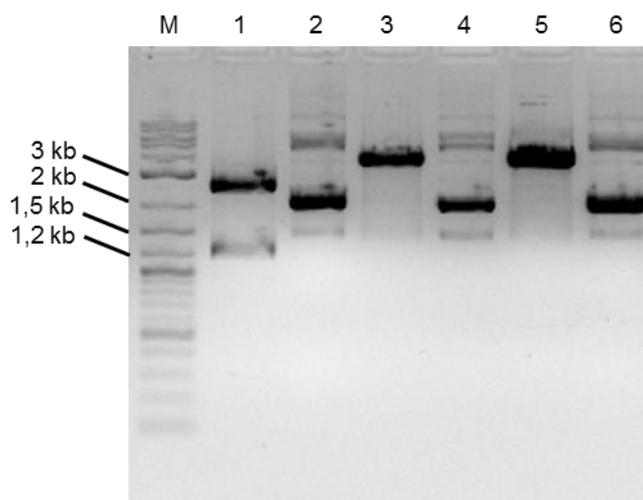


Figura 29. Confirmação da clonagem do fragmento pKpn/pPOLI. Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão com NdeI de plasmídeos extraídos de clones da transformação pBlue-tel8+pKpn/pPOLI. Poços 1, 3, 5: plasmídeos digeridos com NdeI. Poços 2, 4, 6: plasmídeos intactos. M: marcador 2-Log DNA Ladder (New England Biolabs).

A digestão com NdeI deveria liberar um fragmento de ~1200 pb e outro fragmento de ~2300 pb, visto que agora o vetor possuía 2 sítios de NdeI. O clone 1 exibiu o resultado esperado, enquanto os demais clones possuíam apenas um sítio de NdeI e foram linearizados pela digestão. O vetor proveniente do clone 1 foi adicionalmente verificado através da digestão individual com NotI, BamHI e AvrII e foi chamado de pBlue-tel8f. Este plasmídeo foi finalmente usado para a montagem dos braços do cromossomo artificial.

De posse do vetor pBlue-tel8f, foram então iniciadas duas estratégias de clonagem utilizando este vetor: em uma delas, seria clonado o cassete de expressão do gene *ADE3*; na outra, seria clonada a região contendo a marca de seleção *Sh ble* e a sequência ARS1, assim como o centrômero 4.

5.3.4. Braço *ADE3*

Na primeira clonagem, a marca *ADE3* foi obtida a partir da digestão do vetor pPICH-*ADE3* com a enzima *NotI*. Esta digestão liberou os cerca de 4000 pb contendo o gene e seu promotor e terminador, que foram purificados. Este fragmento foi ligado ao vetor pBlue-tel8f que havia sido linearizado com *NotI* e defosforilado. A construção foi confirmada por digestão com a enzima de restrição *SacI* e os resultados são mostrados na figura 30.

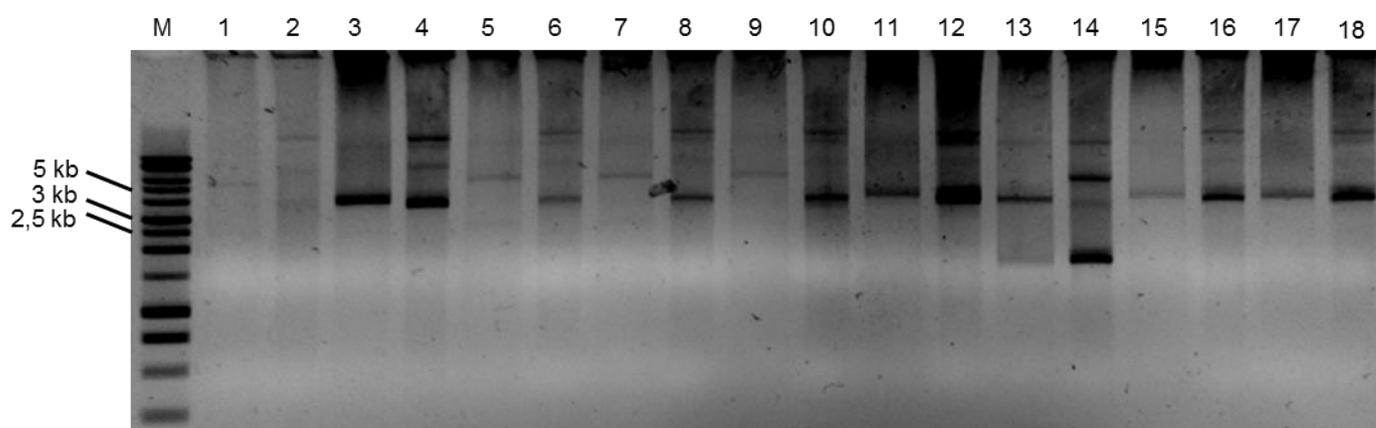


Figura 30. Confirmação da clonagem do cassette *ADE3* no vetor pBlue-tel8f. Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão com *SacI* de plasmídeos extraídos de clones da transformação pBlue-tel8f+*ADE3*. Poços 1, 3, 5, 7, 9, 11, 13, 15 e 17: plasmídeos digeridos com *SacI*. Poços 2, 4, 6, 8, 10, 12, 14, 16 e 18: plasmídeos intactos. M: marcador 1 kb Plus DNA Ladder (Promega).

A partir desta digestão, os vetores com a correta clonagem do gene *ADE3* liberariam bandas de 2428 pb e 5301 pb, o que pode ser visto nos poços 1, 5, 7 e 9 da figura 30. A orientação do cassette de expressão do gene *ADE3* foi escolhida de maneira que o promotor estivesse mais distante do telômero, a fim de evitar o silenciamento da expressão gênica pela posição próxima ao telômero (GOTTSCHLING *et al.*, 1990). A figura 31 ilustra a conformação final do vetor obtido, chamado de pADE-tel8.

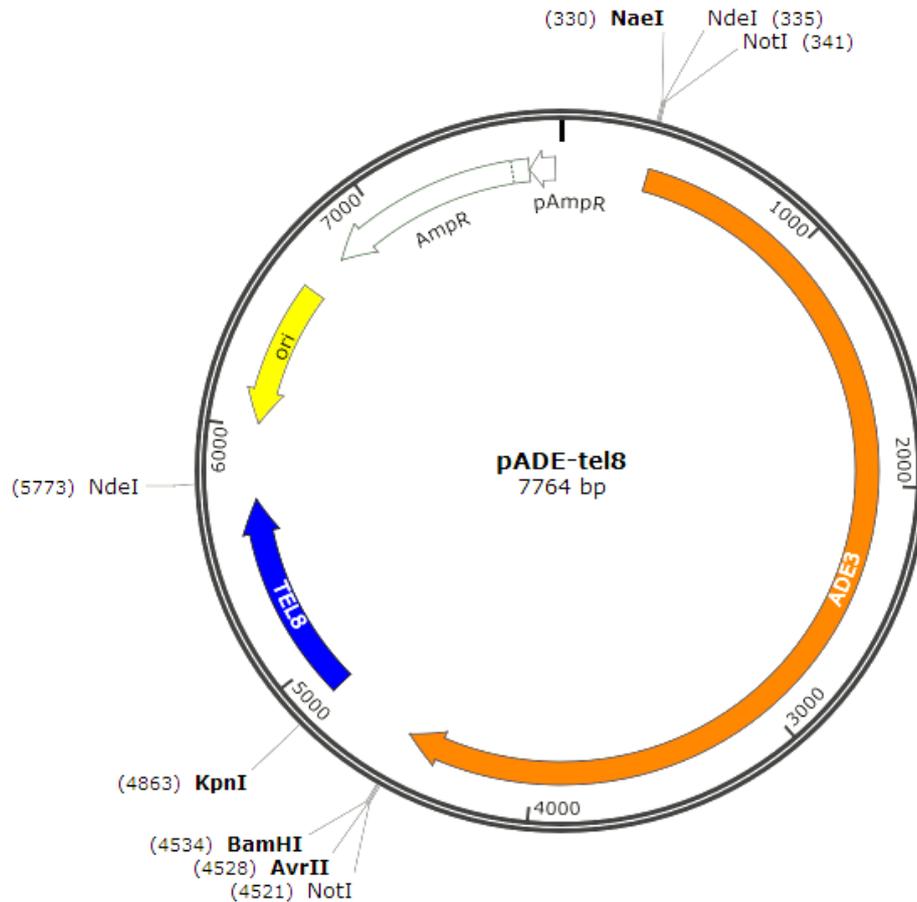


Figura 31. Mapa do vetor pADE-tel8. O cassete *ADE3* é indicado em alaranjado, repetições teloméricas em azul. Em amarelo, origem de replicação e em branco a marca de seleção de *E. coli*. Os sítios de *NdeI* serão utilizados para a liberação do braço linear.

5.3.5. Braço CEN

O vetor pBlue-tel8f foi utilizado também na clonagem do centrômero, da sequência *ARS1* e da outra marca de seleção, o gene *Sh ble* de resistência ao antibiótico zeocina. Uma versão deste vetor também foi usada na clonagem do promotor induzido *AOX1* na direção do centrômero. A primeira sequência clonada neste braço do cromossomo foi o centrômero 4. Este foi liberado do vetor pPICH-CEN4 através da digestão com *BamHI* e purificação, enquanto o vetor pBlue-tel8f foi digerido com a mesma enzima e ambos foram submetidos a ligação. *E. coli* foi transformada com esta ligação e foram realizadas 10 extrações plasmidiais. As extrações que apresentaram maior tamanho de

plasmídeo intacto na primeira inspeção em gel de agarose foram digeridas com a enzima BamHI (figura 32).

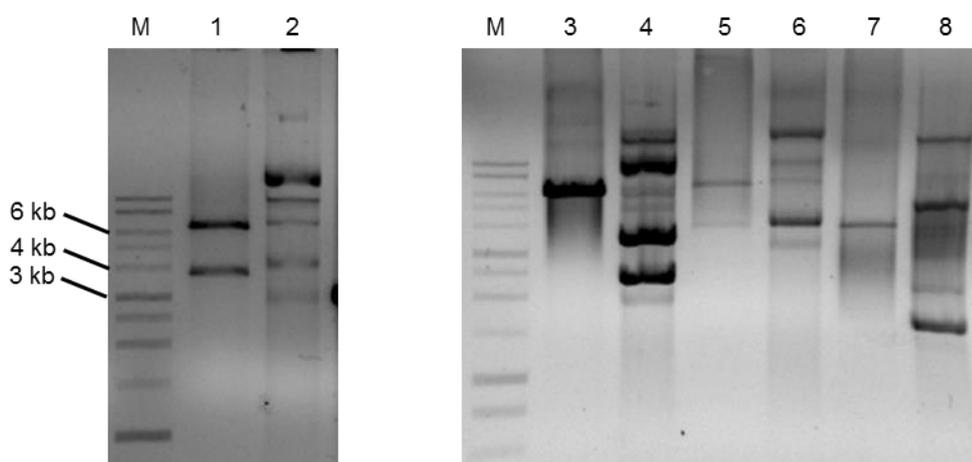


Figura 32. Confirmação da clonagem da sequência CEN4. Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão com BamHI de plasmídeos extraídos de clones da transformação pBlue-tel8f+CEN4. Poços 1, 3, 5 e 7: plasmídeos digeridos com BamHI. Poços 2, 4, 6, 8: plasmídeos intactos. M: marcador 1 kb Plus DNA Ladder (Promega).

No caso de uma clonagem bem-sucedida, seriam observadas bandas de 6596 e 3549 pb nesta digestão. Os poços 1 e 5 da figura 32 apresentaram o resultado desejado, e os clones foram nomeados de pBlue-tel8C e armazenados para as etapas seguintes de clonagem.

A próxima clonagem neste braço do cromossomo artificial foi da região contendo a marca *Sh ble* e a sequência ARS1. Ambas foram clonadas juntamente devido à localização adjacente no vetor pPICH-CEN4. Estas sequências foram amplificadas por meio de PCR com *primers* ARSzeo-F e ARSzeo-R a partir do vetor pPICH-CEN4 e unidas ao vetor pBlue-tel8C previamente digerido com NotI através de recombinação *in vitro* com o *kit* NEBuilder Assembly. A transformação em *E. coli* XL10-gold foi plaqueada em meio LB com ampicilina. Foram feitos 9 inóculos para extração plasmidial utilizando zeocina como antibiótico para seleção, dos quais 8 apresentaram crescimento. Os plasmídeos obtidos foram digeridos com NotI. Esta digestão deveria remover o fragmento recém-clonado de aproximadamente 1400 pb, liberando a banda do vetor de ~10,1 kb. A figura 33 mostra que o clone 1 teve

o comportamento esperado, apesar de exibir uma digestão parcial. Este plasmídeo foi chamado de pCEN-tel8 e sua conformação é ilustrada na figura 34.

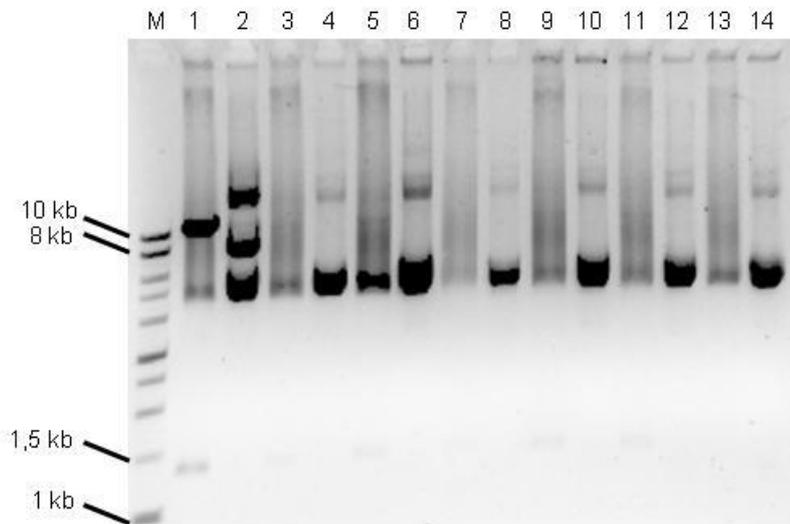


Figura 33. Confirmação da clonagem do fragmento ARSzeo. Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando a digestão com NotI de plasmídeos extraídos de clones da transformação pBlue-tel8fCEN4+ARSzeo. Poços 1, 3, 5, 7, 9, 11 e 13: plasmídeos digeridos com NotI. Poços 2, 4, 6, 8: plasmídeos intactos. M: marcador 1 kb Plus DNA Ladder (Promega).

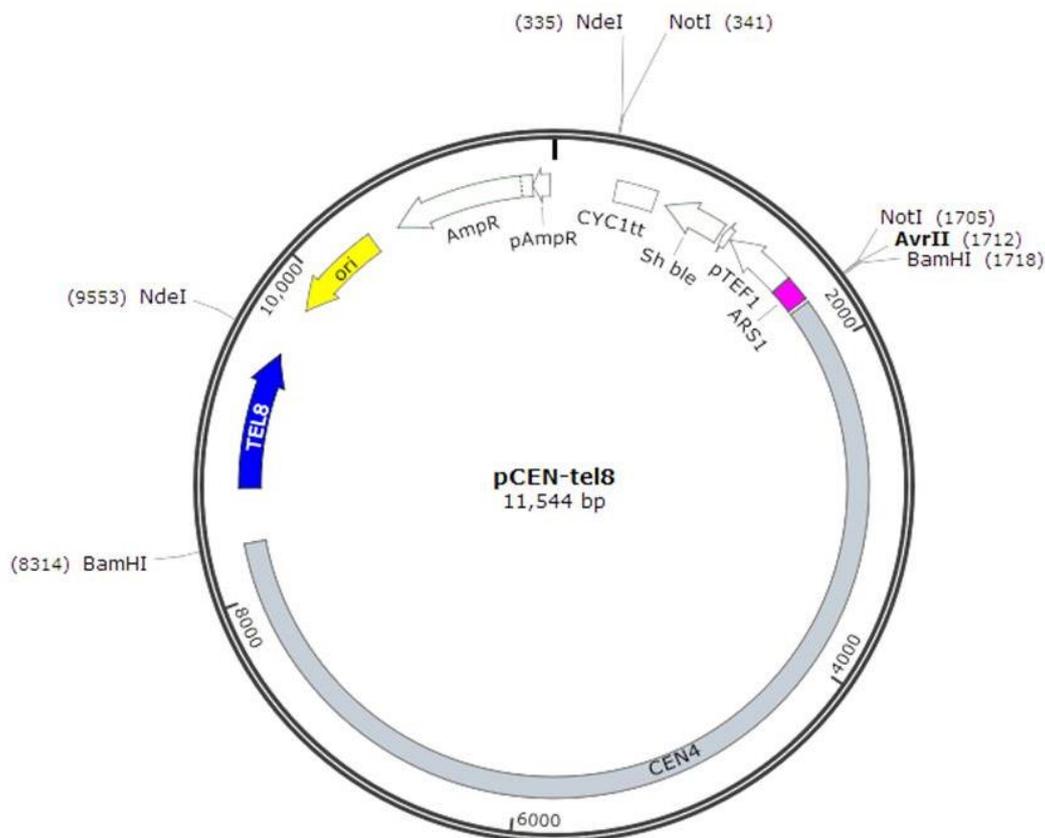


Figura 34. Mapa do vetor pCEN-tel8. O centrômero 4 é indicado em cinza, a sequência telomérica em azul e ARS1 em rosa. Em branco, os cassetes de resistência a zeocina e ampicilina, em amarelo a origem de replicação de *E. coli*. Os sítios de NdeI serão utilizados para a liberação do braço linear.

A última clonagem realizada nesta etapa envolveu o sítio de AvrII presente no vetor pCEN-tel8. Este sítio foi utilizado para a clonagem do promotor induzido de *K. phaffii*, P_{AOX1} . A clonagem teve como objetivo avaliar o controle da função do centrômero pela indução da expressão de sua sequência. Um promotor induzido presente na direção do centrômero em um plasmídeo centromérico de levedura é capaz de interferir na segregação deste vetor durante a divisão celular, criando uma população de células com múltiplas cópias (RESNICK; WESTMORELAND; BLOOM, 1990). No caso do promotor *AOX1*, o crescimento em metanol como fonte de carbono interferiria na estabilidade do cromossomo artificial e o número de cópias do vetor por célula aumentaria. Esta função seria útil no caso da utilização do cromossomo artificial em aplicações de expressão heteróloga e engenharia metabólica nas quais é desejável um aumento da dosagem dos genes inseridos a fim de

aumentar a produção das proteínas e/ou do metabólito de interesse (CHLEBOWICZ-ŚLEDZIEWSKA; ŚLEDZIEWSKI, 1985; LIU *et al.*, 2014).

O promotor P_{AOX1} foi amplificado a partir do genoma de *K. phaffii* X-33 utilizando os *primers* AOX1-F e AOX1-R, que continham sequências para recombinação homóloga com o vetor pCEN-tel8. Este foi digerido com a enzima AvrII, purificado e submetido à recombinação *in vitro* com o fragmento amplificado do promotor $pAOX1$ através do *kit* NEBuilder Assembly. *E. coli* foi transformada com o produto da reação e foram realizadas 7 extrações de plasmídeos. O produto destas extrações foi digerido com AvrII, porém, o fragmento de ~900 pb liberado por esta digestão foi de difícil visualização no gel de agarose devido à diferença em tamanho com o restante do vetor (~11,6 kb). Portanto, foi realizada uma PCR de colônia de confirmação com os *primers* AOX1-F e AOX1-R utilizando 4 dos clones obtidos (figura 35).

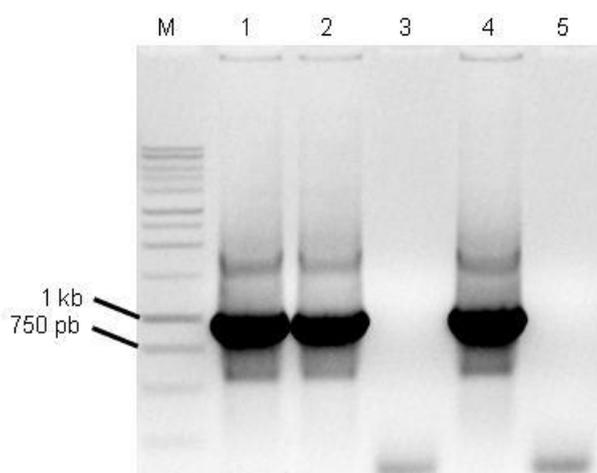


Figura 35. Clonagem do fragmento P_{AOX1} . Eletroforese em gel de agarose 1% corado com brometo de etídeo mostrando as reações de PCR de colônia com *primers* AOX1-F e AOX1-R dos clones da transformação pCEN-tel8+ P_{AOX1} . Poços 1, 2, 3 e 4: clones 1, 2, 3 e 4. Poço 5: controle da reação de PCR. M: marcador 1 kb Plus DNA Ladder (Promega).

Os clones 1, 2 e 4 apresentaram uma banda de aproximadamente 900 pb referente à presença do P_{AOX1} , portanto foram armazenados e o vetor final foi chamado de pCENA-tel8. Este vetor carrega uma versão alternativa do braço que contém o centrômero do cromossomo artificial e foi construído para

verificar se a presença de um promotor induzido interfere na atividade do centrômero. Seu mapa é apresentado na figura 36.

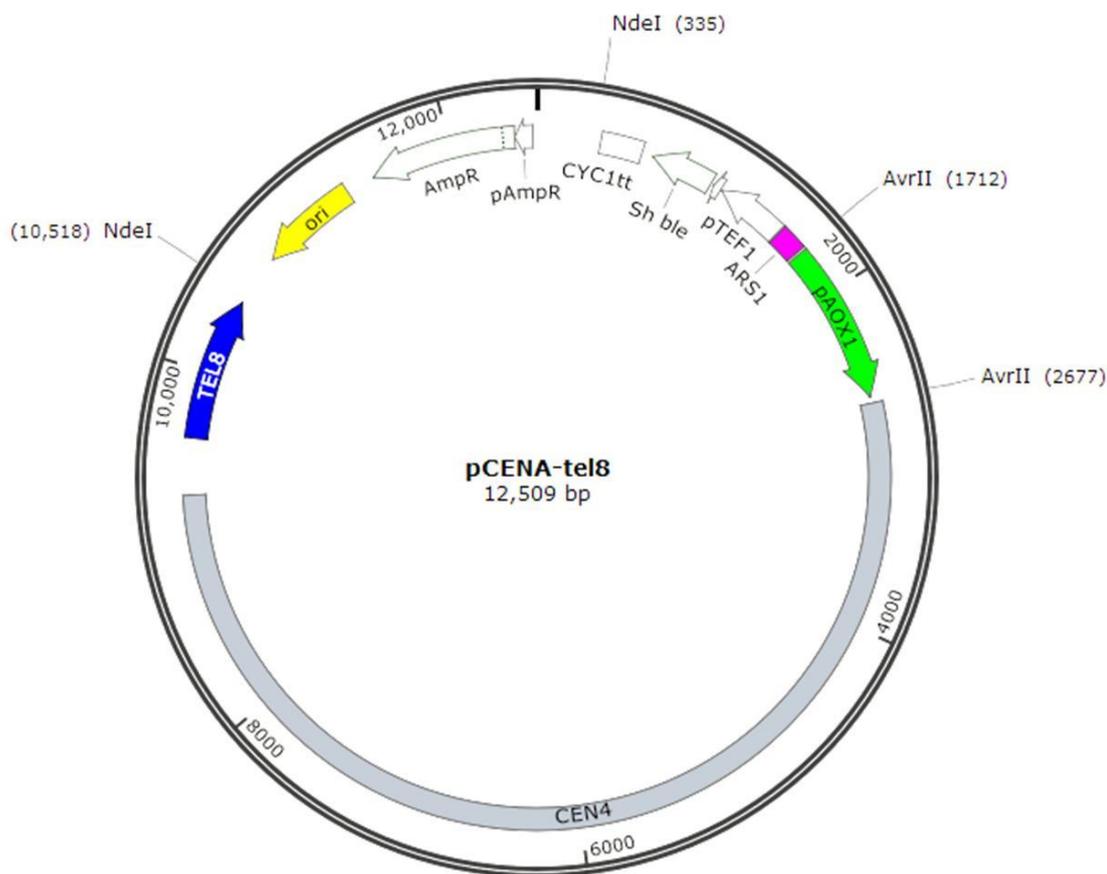


Figura 36. Mapa do vetor pCENA-tel8. O centrômero 4 é indicado em cinza, repetições teloméricas em azul, em rosa a sequência ARS1. P_{AOX1} é indicado em verde. Em branco, cassetes de resistência a ampicilina e zeocina; em amarelo está indicada a origem de replicação de *E. coli*. Os sítios de NdeI serão utilizados para a liberação do braço linear.

Uma vez construídos os vetores pADE-tel8, pCEN-tel8 e pCENA-tel8, todos foram digeridos com NdeI e os braços dos YACs foram purificados para transformação da levedura. O braço proveniente do vetor pADE-tel8 foi utilizado em duas transformações: uma com o braço proveniente do vetor pCEN-tel8, outra com o braço liberado pelo vetor pCENA-tel8.

5.3.6. Transformação e montagem do YAC *in vivo*

A última etapa de clonagem deste trabalho envolveu a montagem do YAC linear diretamente na levedura. A montagem *in vivo* de vetores já foi utilizada em *K. phaffii* tanto na montagem de moléculas de DNA lineares, que posteriormente foram integradas ao genoma da levedura, quanto na montagem de vetores circulares (CAMATTARI *et al.*, 2016; YU *et al.*, 2012).

Em um primeiro momento, foram utilizados como insertos dois genes que facilitariam a verificação dos resultados da montagem. Posteriormente, a transformação e montagem do YAC seriam realizadas com grandes insertos, que são o objetivo da construção deste vetor. Os genes inicialmente utilizados foram *EGFP*, gene repórter da proteína fluorescente verde, e *amdS*, marca de seleção que permite o crescimento da levedura em acetamida como fonte de nitrogênio.

O cassete de expressão do gene *EGFP* foi amplificado a partir do vetor pkGFP-Id (BETANCUR *et al.*, 2017) com os *primers* GFP-ADE3 e GFP-*amdS*. Já a amplificação do cassete de expressão do gene *amdS* foi feita a partir do vetor pAMDS (PIVA *et al.*, 2018) com os *primers* *amdS*-GFP e *amdS*-ARS. Ambos os cassetes de expressão possuíam sequências homólogas para recombinação entre si e com os braços lineares do YAC. Os mapas dos vetores usados como molde estão disponíveis nos anexos 6 e 7.

O cassete *EGFP* possui 1879 pb e o cassete *amdS* possui 2602 pb. A região de recombinação entre o cassete *EGFP* e o braço *ADE* do YAC é de 95 pb, enquanto a região para recombinação entre o cassete *amdS* e o braço *CEN* é de 92 pb. Os dois cassetes de expressão possuem 140 pb de homologia para recombinação entre si.

A estratégia de montagem *in vivo* do YAC está ilustrada na figura 37. Os vetores pADE-*tel8* e pCEN-*tel8* foram digeridos com *NdeI* e os braços lineares foram purificados. O vetor pCENA-*tel8* também passou pelo mesmo processo. Os fragmentos amplificados por PCR foram purificados e a linhagem LA3 passou por dois procedimentos de transformação, cada um com uma combinação dos 4 fragmentos (tabela 5). O protocolo seguido foi o de Wu e Letchworth (2004), com uma proporção de 1:2 entre os braços do vetor e os insertos.

Tabela 5. Combinação de fragmentos lineares para transformação de *K. phaffii*.

Fragmentos	Vetor final
Braço ADE + Braço CEN + <i>amdS</i> + <i>EGFP</i>	YAC
Braço ADE + Braço CEN4 + <i>amdS</i> + <i>EGFP</i>	YAC-AOX1

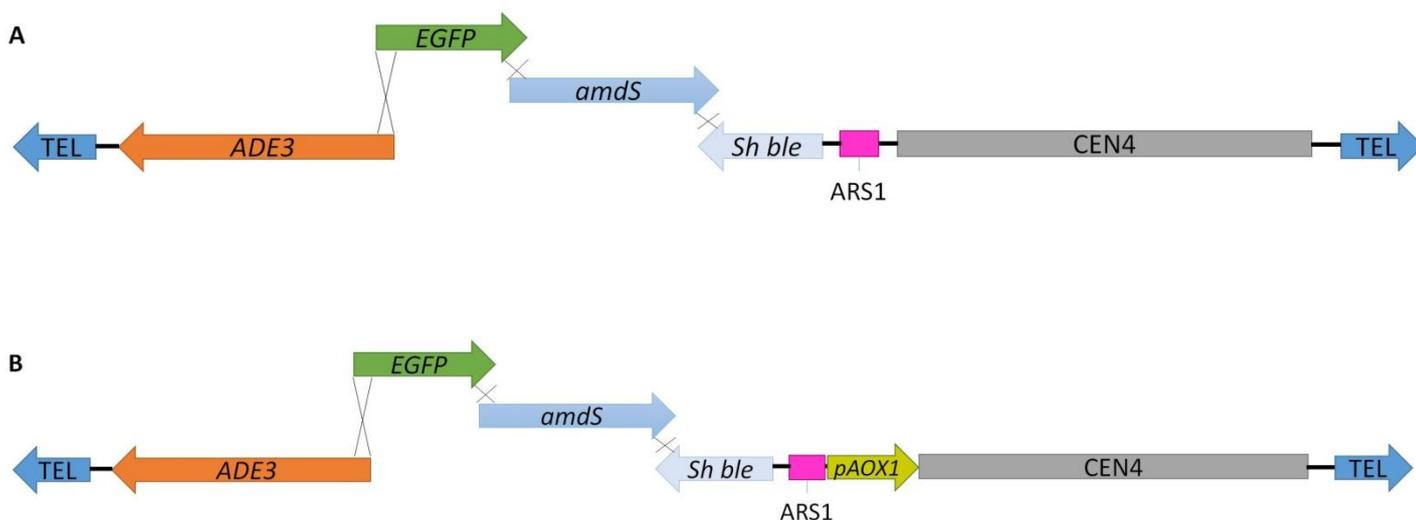


Figura 37. Representação esquemática da estratégia de montagem *in vivo* do cromossomo artificial. A: YAC. B: YAC-AOX1.

A construção final do cromossomo artificial YAC deveria possuir 18,8 kb, enquanto a versão YAC-AOX1 deveria conter 19,6 kb. Após as transformações, a levedura foi plaqueada em meio MD-Ac suplementado de adenina. Este meio de cultura selecionou clones que receberam o cassete do gene *amdS* e eram capazes de utilizar acetamida como fonte de nitrogênio, assim como clones que possuíam o gene *ADE3* e, portanto, eram prototróficos para histidina.

Foram obtidos 2 clones para a construção YAC e 1 clone para a construção YAC-AOX1. Estes foram transferidos para uma placa de meio YPD com zeocina, para verificar a presença desta segunda marca de seleção. Nesta etapa, apenas um clone de cada construção foi capaz de crescer. Os clones YAC e YAC-AOX1 foram então analisados quanto à presença da proteína fluorescente EGFP por meio da extração de proteínas e análise em gel de

poliacrilamida não-desnaturante. A figura 38 ilustra a fluorescência exibida pelo clone YAC crescido em YPD com e sem zeocina.



Figura 38. Confirmação da expressão de EGFP no clone YAC. Gel de poliacrilamida não-desnaturante. Extrato proteico do clone YAC crescido em YPD com zeocina (1) e sem zeocina (2). 3: Linhagem LA3.

O clone YAC foi analisado tanto em meio seletivo quanto não seletivo porque havia a possibilidade deste vetor ser instável na levedura, visto que cromossomos artificiais de tamanhos reduzidos (também chamados de minicromossomos) têm uma menor capacidade de manter-se de maneira estável ao longo das divisões celulares (HIETER *et al.*, 1985). Em ambas as condições de crescimento houve detecção da fluorescência da proteína EGFP, mostrando que o clone YAC era estável na levedura mesmo em condições não seletivas.

Não foi possível detectar fluorescência a partir do clone YAC-AOX1. Este resultado pode significar a ausência do cassete de expressão de GFP, que seria verificada por PCR mais adiante. Outra hipótese seria que a expressão malsucedida se deve à confirmação do DNA adjacente, visto que as sequências centroméricas e teloméricas encontram-se próximas ao cassete e são capazes de prejudicar sua expressão (RANDO; WINSTON, 2012). Este resultado pode se dever ao dobramento inadequado, instabilidade da proteína ou até mesmo ao estresse celular (DAMASCENO; HUANG; BATT, 2012).

5.3.7. Verificação da integridade e estrutura dos YACs

A partir da obtenção dos clones YAC e YAC-AOX1, foram realizadas reações de PCR de colônia para verificar a presença de todos os quatro fragmentos de DNA utilizados na montagem dos cromossomos artificiais. Não foi possível amplificar o cassete *ADE3* completo com os *primers* ADE3upF e ADE3dwR, nem o cassete *amdS* com os *primers* amdS-GFP e amdS-ARS. Em vista destes resultados, foram feitas reações que amplificaram pequenos fragmentos de cada um dos braços do cromossomo artificial, assim como fragmentos dos genes *EGFP* e *amdS* (reações número 1, 2, 3, e 5 – tabela 6). Além disso, foram feitas também amplificações por PCR com *primers* que se anelavam um em cada fragmento utilizado para recombinação *in vivo*, ou seja, que amplificassem as regiões utilizadas na junção do YAC e YAC-AOX1 (reações 4, 6 e 7). Uma PCR adicional foi realizada com *primers* que se anelavam nas sequências do vetor e amplificaram a região contendo o centrômero 4 (reação 8). Os pares de *primers* utilizados e os fragmentos esperados estão relacionados na tabela 6.

Tabela 6. *Primers* utilizados nas PCRs de confirmação dos clones YAC e YAC-AOX1.

Reação de PCR	Par de <i>primers</i>	Tamanho do fragmento (pb)
1	5zeo e 3zeo	385
2	qAMS-F e qAMS-R	140
3	GFP-amdS e GFP-ADE3	1879
4	c-amdSGFP-F e c-amdSGFP-R	598
5	pPOLI e pKpn	365
6	c-ADEGFP-F e c-ADEGFP-R	459
7	qAMS-F e 5zeo	1644
8	pKpn e 3zeo	7986 (YAC) 8951 (YAC-AOX1)

A figura 39 ilustra a representação gráfica da conformação final do YAC e os fragmentos amplificados em cada PCR estão indicados por números. As regiões indicadas em vermelho são as intersecções entre os fragmentos unidos por recombinação *in vivo*.

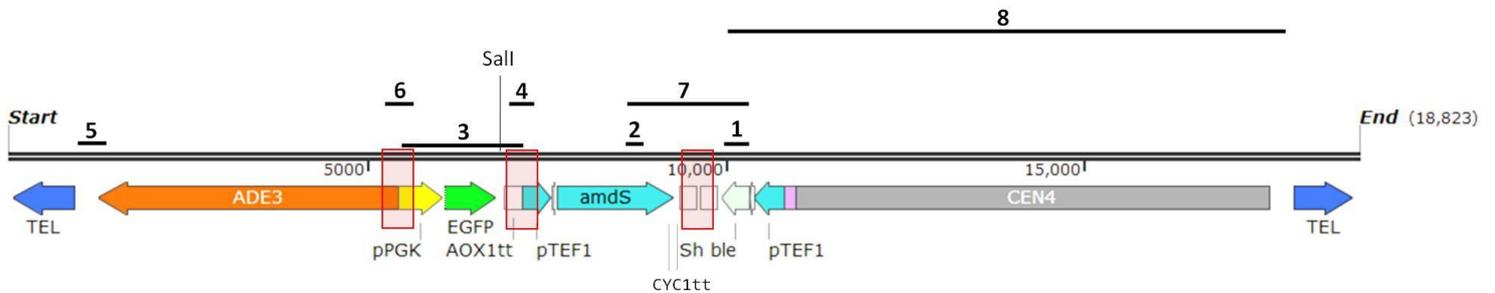


Figura 39. Diagrama representativo da construção do YAC final. Linhas pretas numeradas representam PCRs de confirmação. Em vermelho, regiões usadas para recombinação *in vivo*.

Os resultados obtidos mostram que a montagem dos fragmentos por recombinação *in vivo* foi bem-sucedida, ou seja, os quatro fragmentos encontravam-se organizados de maneira contígua dentro das células tanto do clone YAC (figura 40) quanto do clone YAC-AOX1 (figura 41).

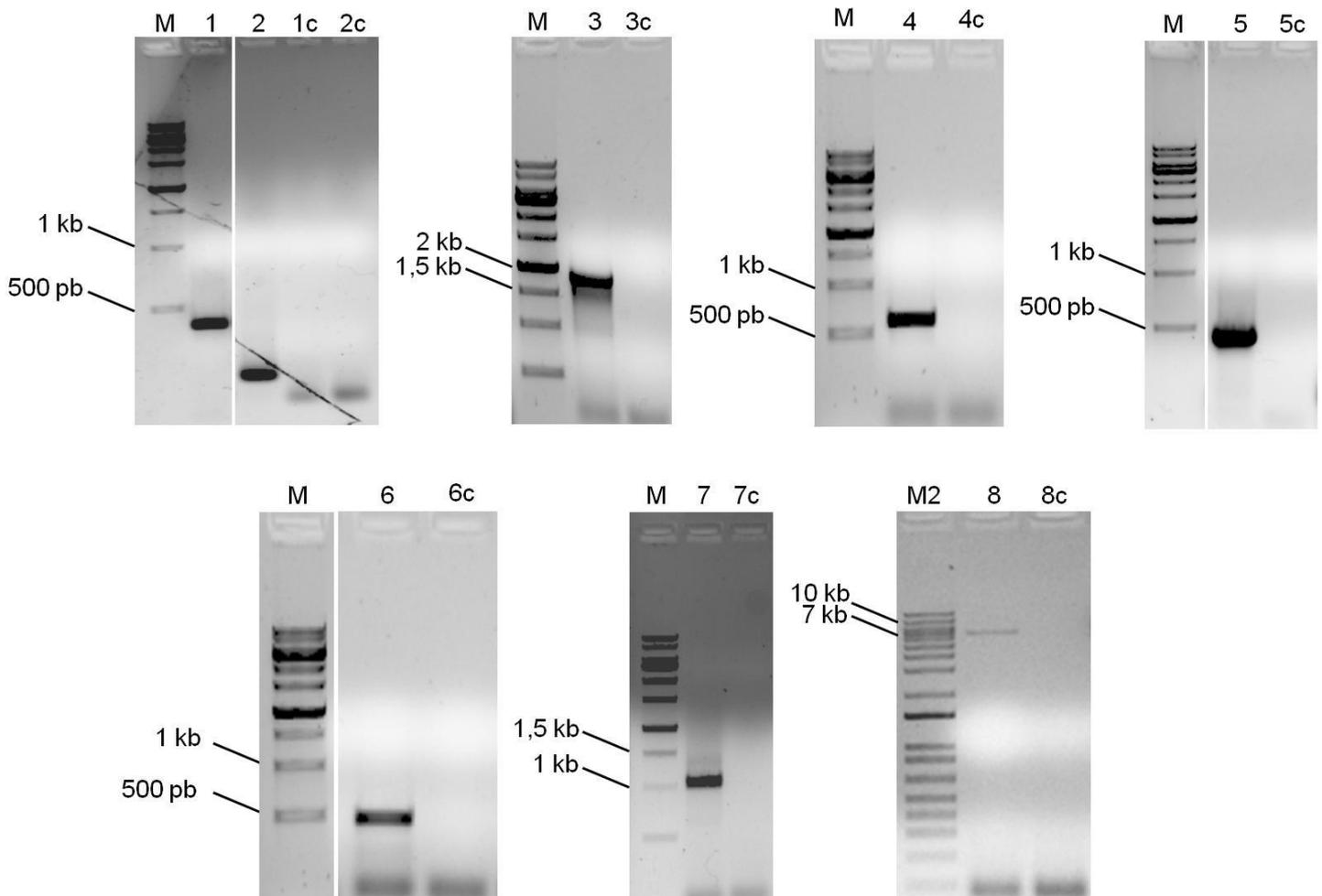


Figura 40. PCRs de confirmação da montagem do vetor YAC. Eletroforeses em géis de agarose 1% corados com brometo de etídeo exibindo PCRs de colônia de confirmação do clone YAC. 1-8: reações 1 a 8 de acordo com tabela 6. 1c-8c: controles negativos das PCRs. M: marcador 1 kb Ladder (Sinapse Inc), M2: marcador 1 kb Plus DNA Ladder (Thermo Fischer Scientific).

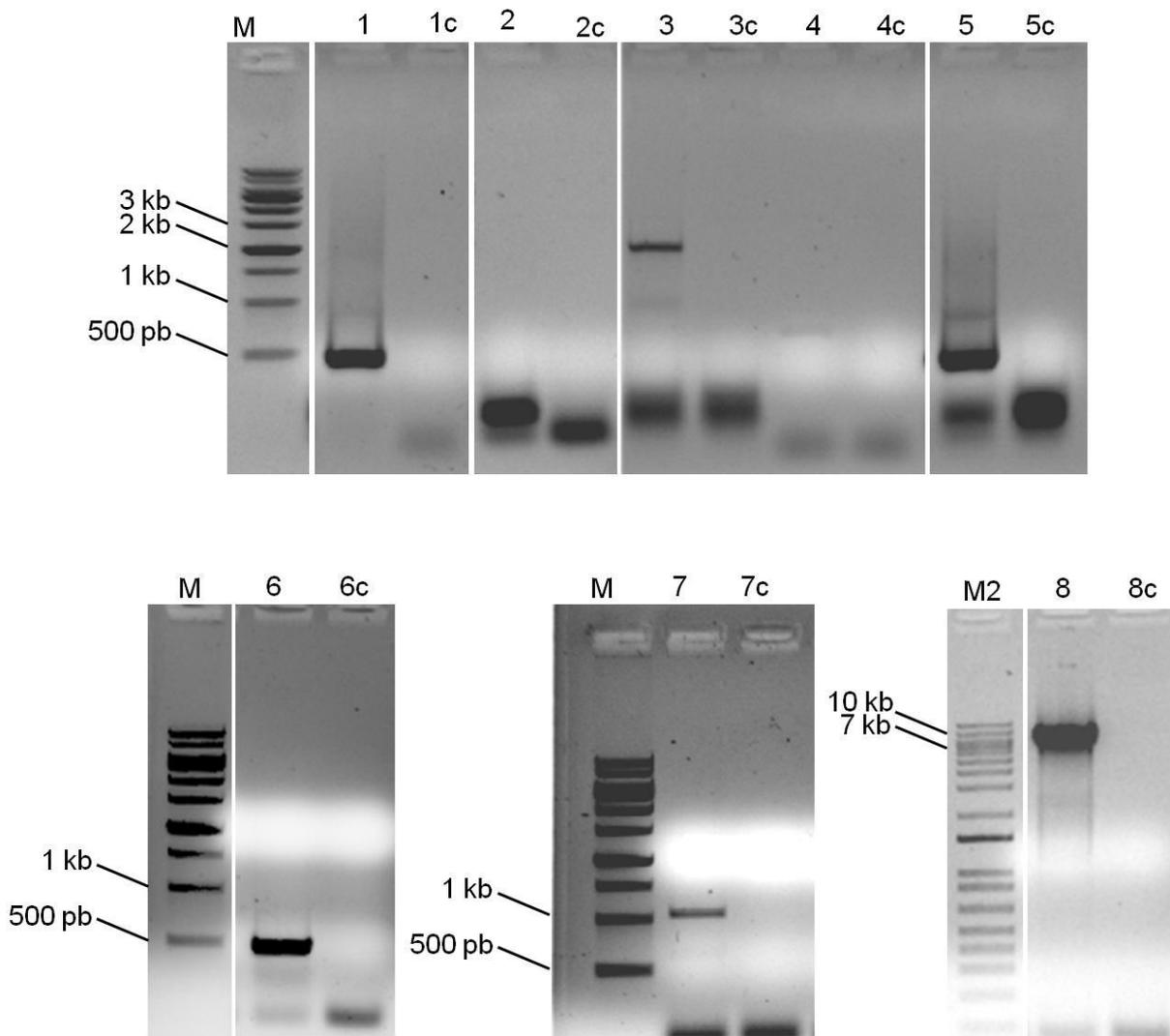


Figura 41. PCRs de confirmação da montagem do vetor YAC-AOX1. Eletroforeses em géis de agarose 1% corados com brometo de etídeo exibindo PCRs de colônia de confirmação do clone YAC-AOX1. 1-8: reações 1 a 8 de acordo com tabela 6. 1c-8c: controles negativos das PCRs. M: marcador 1 kb Ladder (Sinapse Inc), M2: marcador 1 kb Plus DNA Ladder (Thermo Fischer Scientific).

É importante destacar que, enquanto o tamanho esperado da região amplificada pela reação número 7 era de 1644 pb, em ambos os clones o resultado foi uma banda de pouco mais de 1000 pb. Estes fragmentos amplificados foram purificados e enviados para sequenciamento com os *primers* utilizados na reação 7. O resultado obtido para o clone YAC foi comparado com a sequência teórica e está disponibilizado no anexo 2, porém o sequenciamento do clone YAC-AOX1 não teve resultados satisfatórios. O resultado do clone YAC apresentou algumas discrepâncias entre as reações de

sequenciamento com cada *primer*, como resultado de erros no sequenciamento. Entretanto, ambas as reações indicaram que os dois terminadores *CYC1* dos cassetes de expressão dos genes *amdS* e *Sh ble* foram perdidos, provavelmente por recombinação homóloga. Esta excisão provavelmente ocorreu por recombinação homóloga, visto que os terminadores se localizavam próximos às extremidades usadas para a montagem do vetor. Em construções heterólogas utilizadas em leveduras, sequências repetitivas podem causar eventos de recombinação intramolecular que excisam cassetes de expressão inteiros (AW; POLIZZI, 2013). Este tipo de recombinação também é o responsável pela baixa estabilidade de clones com múltiplas cópias de genes heterólogos em *K. phaffii* (ZHU *et al.*, 2009). A perda dos terminadores pode influenciar a quantidade e a estabilidade do mRNA do gene de interesse (REDDEN; MORSE; ALPER, 2014), no entanto a expressão tanto de EGFP quanto de acetamidase foi detectada no clone YAC.

Um experimento de eletroforese de campo pulsado (PFGE) foi realizado a fim de verificar se os vetores YAC e YAC-AOX1 seriam visualizáveis neste tipo de gel, em comparação com os cromossomos nativos da levedura. Os vetores deveriam ter uma conformação linear e não integrada ao genoma de *K. phaffii*. Primeiramente foi feita uma corrida curta, de 6 horas, para separar moléculas de DNA pequenas e verificar a existência dos YACs (figura 42).

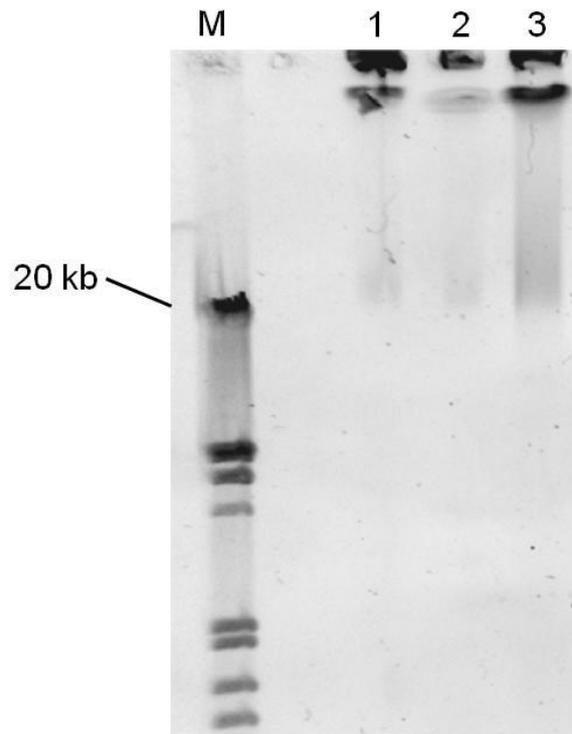


Figura 42. Corrida curta de PFGE. Gel de agarose 1% corado com brometo de etídeo. 1: linhagem LA3. 2: linhagem transformada com YAC. 3: linhagem transformada com YAC-AOX1. M: marcador 2-Log DNA Ladder (New England Biolabs).

Este primeiro resultado revelou que os cromossomos das leveduras não foram separados, sendo acumulados em uma forte banda no topo do gel, e que uma quantidade considerável de material ficou presa aos cubos de amostra. Também se pode observar uma banda com baixa definição na região correspondente aos 20 kb em todas as amostras, inclusive a linhagem LA3 sem transformação. O tamanho de 20 kb coincide com o de um plasmídeo linear descrito em *K. phaffii* CBS 7435 (KUBERL *et al.*, 2011). Este plasmídeo carrega genes homólogos aos plasmídeos *killer* presentes em leveduras como *Kluyveromyces lactis* e *S. cerevisiae*. Apesar de não ter sido detectado em *K. phaffii* GS115, que seria a linhagem mais próxima da linhagem LA3 utilizada neste trabalho, o plasmídeo pode estar presente em GS115 com um menor número de cópias. Transcritos correspondentes aos genes localizados neste plasmídeo foram detectados por RNA-seq, porém em quantidades reduzidas (LOVE *et al.*, 2016). Plasmídeos lineares de 13,1 e 9,5 kb também foram

descritos em linhagens selvagens de *K. phaffii*, expressando os mesmos genes relacionados aos plasmídeos *killer* de *K. lactis* (STURMBERGER *et al.*, 2016).

A partir deste gel, foi realizado um experimento de Southern Blot utilizando a sonda *Sh ble* para localizar os vetores YAC e YAC-AOX1, porém não foi observada quimioluminescência nem na região correspondente aos cromossomos artificiais, nem na região do DNA genômico da levedura.

A segunda condição utilizada para as corridas de PFGE foi de 48 horas e separou os cromossomos nativos da levedura, como mostrado na figura 43. Desta vez, a linhagem X-33 foi utilizada como controle além da linhagem LA3.

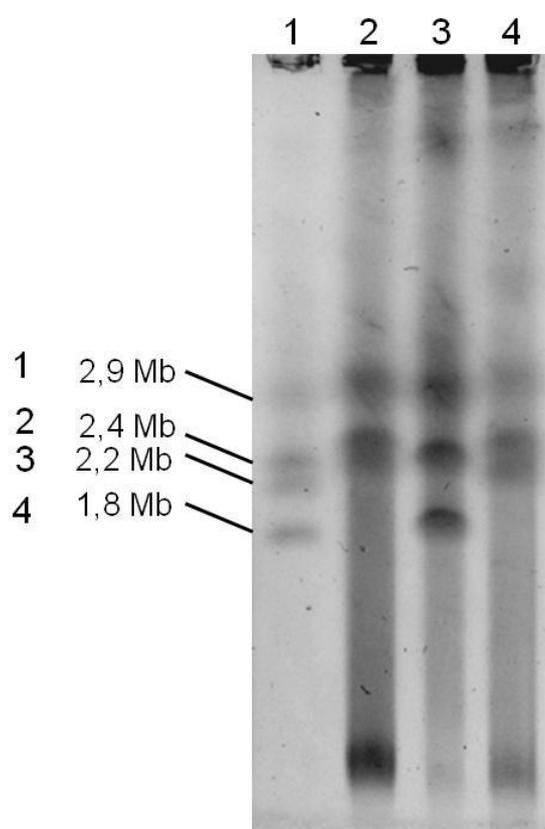


Figura 43. Corrida longa de PFGE. Gel de agarose 1% corado com brometo de etídeo. 1: linhagem X-33. 2: linhagem LA3. 3: linhagem LA3 transformada com YAC. 4: linhagem LA3 transformada com YAC-AOX1.

Os cromossomos nativos de *K. phaffii* possuem aproximadamente 1,8, 2,2, 2,4 e 2,9 Mb (DE SCHUTTER *et al.*, 2009) e são ilustrados na figura 43 na amostra da linhagem X-33. Considerando que esta linhagem possui os tamanhos de cromossomos correspondentes aos publicados na literatura, a

linhagem LA3 apresenta uma discrepância mesmo antes da transformação com o cromossomo artificial. Esta linhagem, que possui deleções nos genes *AMD2*, *ADE2* e *ADE3*, apresenta um cromossomo menor que os demais e não apresenta a banda correspondente ao cromossomo de 1,8 kb. Este é o cromossomo 4, onde se localiza a sequência do gene *ADE3*. Os demais genes deletados nesta linhagem encontram-se no cromossomo 3, com 2,2 Mpb de tamanho, que apresenta difícil visualização nesta amostra da linhagem LA3. As deleções dos genes removeram alguns kb da sequência destes cromossomos, mas não o suficiente para gerar uma mudança deste tipo no padrão de corrida na PFGE.

Uma hipótese para o que ocorreu com a linhagem LA3 é a de que este rearranjo cromossômico pode ter sido causado pela recombinação entre sítios *loxP* na reciclagem das marcas de seleção empregadas na deleção destes genes. No momento da reciclagem da marca *kan* do locus *ADE3*, a levedura já possuía duas outras sequências *loxP* que restaram como “cicatrices” das reciclagens das marcas no cromossomo 3. Dessa forma, além da excisão da marca de seleção, a recombinase CreA pode ter realizado rearranjos intra e intercromossômicos. As sequências *loxP* e a recombinase CreA já foram utilizadas intencionalmente na geração de variabilidade genômica em cromossomos sintéticos de *S. cerevisiae*. Ao se espalhar sítios *loxP* em locais estratégicos do cromossomo sintético, a expressão da recombinase CreA inseriu deleções, inversões, duplicações e rearranjos complexos na sequência original, criando um conjunto de clones com características únicas (SHEN *et al.*, 2016).

Ainda na figura 43, a amostra correspondente ao clone YAC apresenta um padrão de bandas semelhante ao exibido por X-33, porém as bandas correspondentes aos cromossomos 2 e 3 (2,4 e 2,2 Mb, respectivamente) estão indistinguíveis entre si. A presença do cromossomo artificial, principalmente considerando que o centrômero do cromossomo 4 foi o escolhido na construção do vetor, pode ter contribuído para rearranjos adicionais nos cromossomos da levedura. Já a amostra do clone YAC-AOX1 corresponde ao padrão de rearranjos observado originalmente na linhagem LA3 e não sofreu mudanças adicionais. A presença de sequências homólogas ao genoma da levedura em vetores replicativos pode levar à integração destes plasmídeos por recombinação homóloga, especialmente caso os YACs sejam

pouco estáveis em sua forma linear (MADSEN; VISMANS; BRINCH-PEDERSEN, 2016). Neste caso, os vetores possuem sequências *upstream* e *downstream* da região codante do gene *ADE3*, o centrômero 4 e as repetições teloméricas como potenciais pontos de integração.

Caso o cromossomo artificial tenha se integrado no genoma da levedura, cria-se um cromossomo com dois centrômeros. Em *S. cerevisiae*, este tipo de cromossomo sofre um grande estresse mecânico durante a divisão celular, visto que o fuso mitótico pode puxar uma mesma molécula para pólos opostos da célula. Como resultado, o cromossomo acaba se rompendo (HABER; THORBURN; ROGERS, 1984). A presença de quebras de dupla fita nestes cromossomos dá início a processos de recombinação homóloga e não-homóloga que criam cromossomos com grandes rearranjos, deleções, duplicações e translocações (PENNANEACH; KOLODNER, 2009)

Este processo pode ser evitado caso um dos centrômeros seja inativado ou deletado, como é o caso em *S. pombe*, ou caso a distância entre eles seja pequena de maneira que ambos os pontos de ligação ao fuso mitótico sejam orientados para o mesmo sentido na divisão celular (SATO *et al.*, 2012; STIMPSON; MATHENY; SULLIVAN, 2012). Neste trabalho, a visualização de um padrão cromossomômico no clone YAC diferente daquele observado na linhagem LA3 pode ter ocorrido devido à integração do cromossomo artificial, apesar deste ter sido montado corretamente. Já o clone YAC-AOX1 não parece ter interferido no padrão de bandas em PFGE exibido pela linhagem LA3, o que indica que o vetor pode ter se mantido de maneira linear. Alternativamente, o vetor YAC-AOX1 pode ter se integrado no genoma da levedura, utilizando o próprio *locus* centromérico ou contornado os problemas decorrentes dos cromossomos dicêntricos como ilustrado anteriormente.

Para verificar se houve ou não integração dos cromossomos artificiais no genoma de *K. phaffii*, foi realizado um Southern Blot. Neste experimento, não foi possível obter resultados conclusivos com a linhagem transformada com YAC-AOX1. O DNA total foi extraído da linhagem LA3 transformada com YAC e então digerido com a enzima Sall (posição do sítio indicada na figura 39), que divide o cromossomo artificial em duas partes: uma de 6809 pb e uma de 12014 pb. Estes tamanhos poderiam variar caso os telômeros presentes no YAC fossem processados pela célula. Os telômeros nativos de *K. phaffii*

possuem de 100 a 340 pb de sequências repetitivas, enquanto os telômeros construídos neste trabalho possuíam 800 pb de repetições (LOVE *et al.*, 2016). Considerando também que o resultado da PCR de confirmação número 7 indicou a perda dos terminadores CYC1, a banda de 12 kb acabaria ficando com ~11,5 kb.

Caso o cromossomo artificial estivesse integrado ao genoma, os resultados seriam diferentes dos esperados. Este ensaio utilizou duas sondas diferentes: uma que hibridizou no gene *Sh ble* e deveria evidenciar a banda de 11,5 kb, e outra que se anelou ao gene *ADE3* indicando a banda de 6809 pb.

Pode-se observar que existem bandas inespecíficas de tamanhos pequenos, que não foram consideradas nestes resultados. Os tamanhos revelados não correspondem aos tamanhos previstos pela construção do YAC. No caso da hibridização com a sonda ADE (figura 44C), o tamanho revelado pelo Southern Blot foi de ~8 kb. Este tamanho é semelhante ao previsto, de quase 7 kb, portanto poderia ser considerado como efeito do processamento *in vivo* dos telômeros. Entretanto, a membrana hibridizada com a sonda *Sh ble* (figura 44B) revelou uma banda de pouco mais de 4 kb, muito aquém do tamanho esperado de 12 kb. Este tamanho de 4 kb corresponde à distância entre o sítio de *Sma*I e o início da sequência centromérica, como se esta tivesse sido perdida.

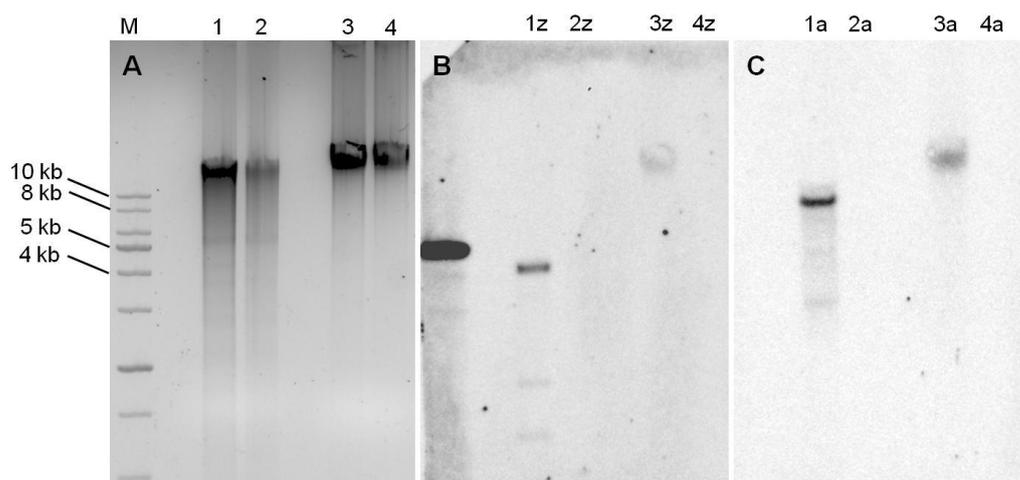


Figura 44. Southern Blot do clone YAC. **A:** Eletroforese em gel de agarose 1% corado com brometo de etídeo. M: marcador 1 Kb Ladder (Sinapse Inc). 1: DNA total do clone YAC digerido com a enzima *Sall*. 2: DNA total da linhagem LA3 digerido com *Sall*. 3: DNA total do clone YAC. 4: DNA total da linhagem LA3. **B:** Southern Blot hibridizado com a sonda *Sh ble*. **C:** Southern Blot hibridizado com a sonda *ADE3*.

Caso tenha ocorrido uma integração no genoma, este tamanho indica que o centrômero deve ter sido excisado para manter a estabilidade do cromossomo e a integração deve ter ocorrido em um local próximo a um sítio de SmaI, para que o fragmento tenha sido liberado com pouco mais de 4 kb. Considerando que o sítio de restrição de SmaI possui 6 pb, estatisticamente este tipo de sítio deve ser encontrado uma vez a cada 4096 (46) pb. É um sítio de restrição comumente encontrado no genoma, portanto poderia contribuir para o resultado obtido no experimento de Southern Blot. Dessa forma, o YAC poderia estar integrado ao genoma, explicando o resultado de bandas diferente do esperado tanto para a sonda ADE quanto para a sonda Sh ble.

Entretanto, o resultado obtido pela PCR número 8 (figuras 40 e 41) indica que o centrômero 4 estaria clonado entre as sequências do YAC nesta mesma linhagem e contradiz a hipótese de perda do centrômero levantada pelo Southern Blot. Como o efeito de excisão da sequência centromérica pode ocorrer independentemente em cada célula ao longo do crescimento da linhagem, no momento da extração de DNA pode-se estar trabalhando com uma população mista de células, consistindo em uma maioria que possui o YAC integrado ao genoma e uma minoria que o possui em sua forma íntegra. Como o limite de detecção de uma reação de PCR é menor do que o de um experimento de Southern Blot, com relação à quantidade de DNA, o resultado deste segundo experimento mostrou apenas a conformação dominante presente na população.

A integração do YAC pode ter ocorrido como consequência de sua instabilidade mitótica. Esta, por sua vez, pode decorrer de seu tamanho, visto que uma molécula de 20 kb ainda é muitas vezes menor que os cromossomos nativos de 2 Mb. Uma particularidade dos YACs lineares é sua estabilidade proporcional ao tamanho da molécula: conforme o tamanho dos insertos clonados neste tipo de vetor aumenta, sua estabilidade mitótica também aumenta (HIETER *et al.*, 1985).

6. Conclusões e perspectivas

Este trabalho teve como objetivo a construção de linhagens auxotróficas para *K. phaffii* e de um vetor linear capaz de receber insertos de tamanhos e complexidades superiores aos hoje comportados pelos sistemas de expressão disponíveis para a levedura. Através das construções realizadas neste trabalho, foi possível estudar novos métodos de manipulação genética desta levedura, como o efeito de centrômeros em vetores circulares e lineares e os mecanismos de recombinação que interferem nestes processos.

A construção de linhagens com dupla auxotrofia neste trabalho produziu 3 linhagens: LA3, auxotrófica para adenina e histidina, LAU, para adenina e uracila e LAL-Kan, para adenina e leucina. Estas linhagens abrem caminhos para novos vetores e estratégias de transformação, e o sistema *ADE2/ADE3* se mostra uma ferramenta importante no estudo da estabilidade de construções em *K. phaffii*, assim como já o é em *S. cerevisiae*.

O uso de vetores centroméricos representa uma novidade no conjunto de ferramentas disponíveis para a levedura e no estudo funcional de seus centrômeros. A presença de ARS nas regiões centroméricas ainda pode ser estudada com base neste tipo de vetor. Novos trabalhos devem estabelecer regiões-chave destes componentes que possam contribuir para a construção de novos vetores com sequências sintéticas, reduzindo as chances de integração na levedura.

A montagem final dos cromossomos artificiais foi realizada com sucesso utilizando a recombinação *in vivo*, entretanto a conformação adotada por estas sequências na levedura e ao longo das gerações de seu crescimento ainda não foi completamente esclarecida.

Além do artigo publicado com parte dos resultados da obtenção das linhagens auxotróficas de *K. phaffii*, foi publicado um artigo de revisão na revista *Bioengineered* com o título "Molecular strategies to increase the levels of heterologous transcripts in *Komagataella phaffii* for protein production". O artigo completo está disponível no anexo 10.

A fim de dar continuidade ao trabalho desenvolvido, temos a intenção de finalizar a análise dos clones que receberam o cromossomo artificial, principalmente o Southern Blot da linhagem transformada com YAC-AOX1, e

realizar estudos da segregação destes vetores. A construção de YACs com insertos maiores ou com diferentes estruturas teloméricas poderá reduzir sua instabilidade e garantir que sua conformação linear se mantenha na levedura. Finalmente, uma via metabólica de produção de beta-caroteno, com 5 genes, será utilizada como inserto na montagem do YAC de *K. phaffii* e avaliará a capacidade do vetor de ser utilizado em estratégias de engenharia metabólica.

Cromossomos artificiais circulares também serão desenvolvidos e comparados aos sistemas lineares, a fim de se obter uma ferramenta que apresente baixas chances de integração ao genoma. Vetores circulares que possuem a capacidade de se processar e gerar moléculas lineares uma vez transformados podem representar uma alternativa contra a integração indesejada de YACs lineares. Sequências nativas de *K. phaffii*, como o gene *ADE3*, devem ser removidas na versão final do vetor e substituídas por marcas de seleção que não possuam homologia ao genoma. O mesmo cuidado deve ser tomado na construção de cassetes de expressão: promotores e terminadores nativos podem ter seus códons otimizados ou serem substituídos, quando possível, por equivalentes de outras espécies.

7. Referências bibliográficas

AHMAD, M.; HIRS, M.; PICHLER, H.; SCHWAB, H. Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production. **Applied Microbiology and Biotechnology**, v. 98, p. 5301–17, 2014.

ARAYA-GARAY, J. M.; FEIJOO-SIOTA, L.; ROSA-DOS-SANTOS, F.; VEIGA-CRESPO, O.; VILLA, T. G. Construction of new *Pichia pastoris* X-33 strains for production of lycopene and β -carotene. **Applied Microbiology and Biotechnology**, v. 93, p. 2483–92, 2012.

AW, R.; POLIZZI, K. M. Can too many copies spoil the broth? **Microbial Cell Factories**, v. 12, n. 128, 2013.

BAILEY, J. E. Toward a science of metabolic engineering. **Science**, v. 252, p. 1668-75, 1991.

BARLOWE, C. K.; APPLING, D. R. Molecular genetic analysis of *Saccharomyces cerevisiae* C1-Tetrahydrofolate synthase mutants reveals a noncatalytic function of the *ADE3* gene product and an additional folate-dependent enzyme. **Molecular and Cellular Biology**, v. 10, n. 11, p. 5679-87, 1990.

BETANCUR, M. O.; REIS, V. C. B.; NICOLA, A. M.; DE MARCO, J. L.; MORAES, L. M. P. de.; TORRES, F. A. G. Multicopy plasmid integration in *Komagataella phaffii* mediated by a defective auxotrophic marker. **Microbial Cell Factories**, v. 16, n. 99, 2017.

BILL, R. **Recombinant protein production in yeast: Methods and Protocols**. New York: Humana Press, doi: 10.1007/978-1-61779-770-5, 2012.

BRUSCHI, C. V.; GJURACIC, K.; TOSATO, V. **eLS: Yeast Artificial Chromosomes**. Chichester, John Wiley & Sons, 2012.

BURKE, D. T.; CARLE, G. F.; OLSON, M. V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. **Science**, v. 236, p. 806–12, 1985.

CAMATTARI, A.; GOH, A.; YIP, L. Y.; TAN, A. H. M.; NG, S. W.; TRAN, A.; LIU, G.; LIACHKO, I.; DUNHAM, M. J.; RANCATI, G. Characterization of a panARS-based episomal vector in the methylotrophic yeast *Pichia pastoris* for recombinant protein production and synthetic biology applications. **Microbial Cell Factories**, v. 15, n.139, 2016.

CAO, M.; GAO, M.; LOPEZ-GARCIA, C. L.; WU, Y.; SEETHARAM, S.; SEVERIN, A. J.; SHAO, Z. Centromeric DNA facilitates nonconventional yeast genetic engineering. **ACS Synthetic Biology**, v. 6, n. 8, 1545-53, 2017.

CARNICER, M.; PIERICK, A. T.; DAM, J. V.; HEIJNEN, J. J.; ALBIOL, J.; VAN GULIK, W.; FERRER, P. Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. **Microbial Cell Factories**, v. 11, n. 83, 2012.

CELIK, E.; CALIK, P. Production of recombinant proteins by yeast cells. **Biotechnology advances**, v. 30, n. 5, p. 1108–18, 2012.

CEREGHINO, J. L.; CREGG, J. M. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. **FEMS Microbiology Reviews**, v. 24, p. 45–66, 2000.

CEREGHINO, G. P. L.; CEREGHINO, J. L.; SUNGA, A. J.; JOHNSON, M. A.; LIM, M.; GLEESON, M. A. G.; CREGG, J. M. New selectable marker / auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. **Gene**, v. 263, p. 159–69, 2001.

CHLEBOWICZ-ŚLEDZIEWSKA, E.; ŚLEDZIEWSKI, A. Z. Construction of multicopy yeast plasmids with regulated centromere function. **Gene**, v. 39, n. 1, p. 25–31, 1985.

CHOI, B.; BOBROWICZ, P.; DAVIDSON, R. C.; HAMILTON, S. R.; KUNG, S. R.; LI, H.; MIELE, R. G.; NETT, J. H.; WILDT, S.; GERNGROSS, T. U. Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. **Proceedings of the National Academy of Sciences**, v. 100, n. 9, p. 5022-7, 2003.

CHURCH, G. M.; ELOWITZ, M. B.; SMOLKE, C. D.; VOIGHT, C. A.; WEISS, R. Realizing the potential of synthetic biology. **Nature Reviews**, v. 15, p. 289-94,

2014.

COUGHLAN, A. Y.; HANSON, S. J.; BYRNE, K. P.; WOLFE, K. H. Centromeres of the yeast *Komagataella phaffii* (*Pichia pastoris*) have a simple inverted-repeat structure. **Genome Biology and Evolution**, v. 8, n. 8, p. 2482–92, 2016.

CREGG, J. M.; BARRINGER, K. J.; HESSLER, A. Y.; MADDEN, K. R. *Pichia pastoris* as a host system for transformations. **Molecular and Cellular Biology**, v. 5, n. 12, p. 3376–85, 1985.

DAMASCENO, L. M.; HUANG, C. J.; BATT, C. A. Protein secretion in *Pichia pastoris* and advances in protein production. **Applied Microbiology and Biotechnology**, v. 93, n. 1, p. 31–9, 2012.

DANI, G. M.; ZAKIAN, V. A. Mitotic and meiotic stability of linear plasmids in yeast. **Proceedings of the National Academy of Sciences**, v. 80, p. 3406-10, 1983.

DE SCHUTTER, K.; LIN, L.; TIELS, P. VAN HECKE, A.; GLINKA, S.; WEBER-LEHMANN, J. ROUZÉ, P.; VAN DE PEER, Y.; CALLEWAERT, N. Genome sequence of the recombinant protein production host *Pichia pastoris*. **Nature Biotechnology**, v. 27, n. 6, p. 561–66, 2009.

DHAMANKAR, H.; PRATHER, K. L. J. Microbial chemical factories: recent advances in pathway engineering for synthesis of value added chemicals. **Current Opinion in Structural Biology**, v. 21, p. 488-94, 2011.

DROCOURT, D.; CALMELS, T.; REYNES, J.; BARON, M.; TIRABY, G. Cassettes of the *Streptoalloteichus hindustanus ble* gene for transformation of lower and higher eukaryotes to phleomycin resistance. **Nucleic Acids Research**, v. 18, n. 13, p. 4009, 1990.

DU, M.; BATTLES, M. B.; NETT, J. H. A color-based stable multi-copy integrant selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as auxotrophic markers. **Bioengineered Bugs**, v. 3, n. 1, p. 32–7, 2012.

DYMOND, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. **Nature**, v. 477, p. 471-7, 2011.

ESPAUX, L.; MENDEZ-PEREZ, D.; LI, R.; KEASLING, J. Synthetic biology for

microbial production of lipid-based biofuels. **Current Opinion in Chemical Biology**, n. 29, p. 58-65, 2015.

ESSANI, K.; GLIEDER, A.; GEIER, M. Combinatorial pathway assembly in yeast. **AIMS Bioengineering**, v. 2, n. 4, p. 423–36, 2015.

GALANIE, S.; THODEY, K.; TRENCHARD, I. J.; INTERRANTE, M. F.; SMOLKE, C. D. Complete biosynthesis of opionds in yeast. **Science**, v. 349, n. 6252, p. 1095-100, 2015.

GEIER, M.; FAULAND, P.; VOGL, T.; GLIEDER, A. Compact multi-enzyme pathways in *P. pastoris*. **Chemical Communications**, v. 51, n. 9, p. 1643-6, 2015.

GIBSON, D. G.; BENDERS, G. A.; ANDREWS-PFANNKOCH, C.; DENISOVA, E. A.; BADEN-TILLSON, H.; ZAVERI, J.; STOCKWELL, T. B.; BROWNLEY, A.; THOMAS, D. W.; ALGIRE, M. A.; MERRYNA, C.; YOUNG, L.; NOSKOV, V. N.; GLASS, J. I.; VENTER, C.; HUTCHISON III, C. A.; SMITH, H. O. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. **Science**, v. 319, n. 5867, p. 1215–20, 2008.

GOTTSCHLING, D. E.; APARICIO, O. M.; BILLINGTON, B. L.; ZAKIAN, V. A. Position Effect at *S. cerevisiae* Telomeres: Reversible Repression of Pol II Transcription. **Cell**, v. 63, p. 751–62, 1990.

GUILLIERMOND, A. *Zygosaccharomyces pastori*, nouvelle espèce de levures copulation hétérogamique. **Bulletin de la Société Mycologique de France**, v. 36, p. 203–11, 1920.

HABER, J. E.; THORBURN, P. C.; ROGERS, D. Meiotic and mitotic behavior of dicentric chromosomes in *Saccharomyces cerevisiae*. **Genetics**, v. 106, N. 2, p. 185–205, 1984.

HANSON, S. J.; BYRNE, K. P.; WOLFE, K. H. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus *Saccharomyces cerevisiae* system. **Proceedings of the National Academy of Sciences**, v. 111, n. 45, p. E4851–8, 2014.

HARTNER, F. S.; RUTH, C.; LANGENEGGER, D.; JOHNSON, S.; HYKA, P.; LIN-CERECHINO, G. P.; LIN-CERECHINO, J.; KOVAR, K.; CREGG, J.;

GLIEDER, A. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. **Nucleic Acids Research**, v. 36, n. 12, 2008.

HE, J.; DENG, J.; ZHENG, Y.; GU, J. A synergistic effect on the production of S-adenosyl-L-methionine in *Pichia pastoris* by knocking in of S-adenosyl-L-methionine synthase and knocking out of cystathionine- β synthase. **Journal of Biotechnology**, v. 126, n. 4, p. 519–27, 2006.

HIETER, P.; MANN, C.; SNYDER, M.; DAVIS, R. Mitotic stability of yeast chromosomes: a colony color assay that measures nondisjunction and chromosome loss. **Cell**, v. 40, n. 2, p. 381–92, 1985a.

HIETER, P.; PRIDMORE, D.; HEGEMANN, J. H.; THOMAS, M.; DAVIS, R. W.; PHILIPPSEN, P. Functional selection of yeast centromeric DNA. **Cell**, v. 42, p. 913–21, 1985b.

HOMMELSHEIM, C. M., FRANTZESKAKIS, L.; HUANG, M.; ÜLKER, B. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. **Scientific Reports**, v. 4, n. 5052, 2014.

HÜRLIMANN, H. C., LALOO, B.; SIMON-KAYSER, B.; SAINT-MARC, C.; COULPIER, F.; LEMOINE, S.; DAIGNAN-FORNIER, B.; PINSON, B. Physiological and toxic effects of purine intermediate 5-amino-4-imidazolecarboxamide ribonucleotide (AICAR) in yeast. **Journal of Biological Chemistry**, v. 286, n. 35, p. 30994–31002, 2011.

IKENO, M.; GRIMES, B.; OKAZAKI, T.; NAKANO, M.; SAITOG, M.; HOSHINO, H.; MCGILL, N. I.; COOKE, H.; MASUMOTO, H. Construction of YAC-based mammalian artificial chromosomes. **Nature Biotechnology**, v. 16, n. 5, p. 431–9, 1998.

ITAKURA, K.; HIROSE, K.; CREA, R.; RIGGS, A. D. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. **Science**, v. 198, p. 1056-63, 1977.

JENSEN, M.; KEASLING, J. Recent applications of synthetic biology tools for yeast metabolic engineering. **FEMS Yeast Research**, v. 15, 2015.

JEONG, E.; SHIM, W. Y.; KIM, J. H. Metabolic engineering of *Pichia pastoris* for production of hyaluronic acid with high molecular weight. **Journal of**

Biotechnology, v. 185, p. 28–36, 2014.

JOSHI, M. U.; PITTMAN, H. K.; HAISCH, C. E.; VERBANAC, K. M. Real-time PCR to determine transgene copy number and to quantitate the biolocalization of adoptively transferred cells from EGFP-transgenic mice. **BioTechniques**, v. 45, p. 247-58, 2008.

KARIM, A. S.; CURRAN, K. A.; ALPER, H. S. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. **FEMS Yeast Research**, v. 13, n.1, p. 107–16, 2013.

KIM, H.; YOO, S. J.; KANG, H. A. Yeast synthetic biology for the production of recombinant therapeutic proteins. **FEMS Yeast Research**, v. 15, n. 1, 2015.

KITADA, K.; YAMAGUCHI, E.; ARISAWA, M. Isolation of a *Candida glabrata* centromere and its use in construction of plasmid vectors. **Gene**, v. 175, p. 105–8, 1996.

KLEIN, J.; HEAL, J. R.; HAMILTON, W. R.; BOUSSEMGHOUNE, T.; TANGE, T. O.; DELEGRANGE, F.; JAESCHKE, G.; HATSCH, A.; HEIM, J. Yeast synthetic biology platform generates novel chemical structures as scaffolds for drug discovery. **ACS Synthetic Biology**, v. 3, n. 5, p. 314–23, 2014.

KLINNER, U.; SCHÄFER, B. Genetic aspects of targeted insertion mutagenesis in yeasts. **FEMS microbiology reviews**, v. 28, n. 2, p. 201–23, 2004.

KOSHLAND, D.; KENT, J. C.; HARTWELL, L. H. Genetic analysis of the mitotic transmission of minichromosomes. **Cell**, v. 40, n. 2, p. 393–403, 1985.

KOUPRINA, N.; LEEM, S.; SOLOMON, G.; LY, A.; KORIABINE, M.; OTSTOT, J.; PAK, E.; DUTRA, A.; ZHAO, S.; BARRETT, J. C.; LARIONOV, V. Segments missing from the draft human genome sequence can be isolated by transformation-associated recombination cloning in yeast. **EMBO Reports**, v. 4, n. 3, p. 257–62, 2003.

KOUPRINA, N.; LARIONOV, V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. **Nature Reviews Genetics**, v. 7, n. 10, p. 805–12, 2006.

KUBERL, A. SCHNEIDER, J.; THALLINGER, G. G.; ANDERL, I.; WIBBERG,

D.; HAJEK, T.; JAENICKE, S.; BRINKOLF, K.; GOESMANN, A.; SZCZEPANOWSKI, R.; PÜHLER, A.; SCHAB, H.; GLIEDER, A.; PICHLER, H. High-quality genome sequence of *Pichia pastoris* CBS7435. **Journal of Biotechnology**, v. 154, p. 312–20, 2011.

KURTZMAN, C. P. Biotechnological strains of *Komagataella (Pichia) pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. **Journal of Industrial Microbiology and Biotechnology**, v. 36, p. 1435–38, 2009.

LARIONOV, V.; KOUPRINA, N.; GRAVES, J.; CHEN, X. N.; KORENBERG, J. R.; RESNICK, M. A. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. **Proceedings of the National Academy of Sciences**, v. 93, n. 1, p. 491–6, 1996.

LEE, C. C.; WILLIAMS, T. G.; WONG, D. W.; ROBERTSON, G. H. An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. **Plasmid**, v. 54, n. 1, p. 80–5, 2005.

LEE, J. W.; KIM, T. W.; JANG, Y.; CHOI, S.; LEE, S. Y. Systems metabolic engineering for chemicals and materials. **Trends in Biotechnology**, v. 29, n. 8, p. 370-8, 2011.

LI, P.; ANUMANTHAN, A.; GAO, X.; ILANGOVAN, K.; SUZARA, V. V.; DÜZGÜNES, N.; RENUGOPALAKRISHNAN, V. Expression of Recombinant Proteins in *Pichia pastoris*. **Applied Biochemistry and Biotechnology**, v. 142, n. 2, p. 105–24, 2007.

LIACHKO, I.; DUNHAM, M. An autonomously replicating sequence for use in a wide range of budding yeasts. **FEMS Yeast Research**, v. 14, n. 2, p. 364-7, 2013.

LIACHKO, I.; YOUNGBLOOD, R. A.; TSUI, K.; BUBB, K. L.; QUEITSCH, C.; RAGHURAMAN, M. K.; NISLOW, C.; BREWER, B. J.; DUNHAM, M. J. GC-rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*. **PLOS Genetics**, v. 10, n. 3, 2014.

LIANG, S.; WANG, B.; PAN, L.; YE, Y.; HE, M.; HAN, S.; ZHENG, S.; WANG, X.; LIN, Y. Comprehensive structural annotation of *Pichia pastoris* transcriptome

and the response to various carbon sources using deep paired-end RNA sequencing. **BMC genomics**, v. 13, n. 738, 2012a.

LIANG, S.; LIN, Y.; YE, Y. Internal ribosome entry site mediates protein synthesis in yeast *Pichia pastoris*. **Biotechnology letters**, v. 34, n. 5, p. 957–64, 2012b.

LIN-CEREGHINO, J.; HASHIMOTO, M. D.; MOY, A.; CASTELO, J.; ORAZEM, C. C.; KUO, P.; XIONG, S.; GANDHI, V.; HATAE, C. T.; CHAN, A.; LIN-CEREGHINO, G. P. Direct selection of *Pichia pastoris* expression strains using new G418 resistance vectors. **Yeast**, v. 25, n. 4, p. 293–9, 2008.

LIU, L.; OTOUPAL, P.; PAN, A.; ALPER, H. S. Increasing expression level and copy number of a *Yarrowia lipolytica* plasmid through regulated centromere function. **FEMS Yeast Research**, v. 14, n. 7, p. 1124–1127, 2014.

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.; LOVE, J. C. Comparative genomics and transcriptomics of *Pichia pastoris*. **BMC Genomics**, v. 17, n. 550, 2016.

LUO, J.; SU, X.; CORMACK, B. P.; BOEKE, J. D. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. **Nature**, v. 560, n. 7718, p. 392-6, 2018.

LYN CHAN, F.; WONG, L. H. Transcription in the maintenance of centromere chromatin identity. **Nucleic Acids Research**, v. 40, n. 22, p. 11178–88, 2012.

MACAULEY-PATRICK, S.; FAZENDA, M. L.; MCNEIL, B.; HARVEY, L. M. Heterologous protein production using the *Pichia pastoris* expression system. **Yeast**, v. 22, p. 249–70, 2005.

MADSEN, C. K.; VISMANS, G.; BRINCH-PEDERSEN, H. The PARS sequence increase the efficiency of stable *Pichia pastoris* transformation. **Journal of Microbiological Methods**, v. 129, 2016.

MARDIS, E. R. DNA sequencing technologies: 2006-2016. **Nature Protocols**, v. 12, n. 2, p. 213-8, 2017.

MARX, H.; MATTANOVICH, D.; SAUER, M. Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*. **Microbial Cell Factories**, v. 7, n. 23,

2008.

MATTANOVICH, D.; GRAF, A.; STADLMANN, J.; DRAGOSITS, M.; REDL, A.; MAURER, M.; KLEINHEINZ, M.; SAUER, M.; ALTMANN, F.; GASSER, B. Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. **Microbial Cell Factories**, v. 8, n. 29, 2009.

MCPHERSON, J. D.; MARRA, M.; HILLIER, L. *et al.* A physical map of the human genome. **Nature**, v. 409, n. 6822, p. 934–941, 2001.

MONACO, A. P.; LARIN, Z. YACs, BACs, PACs and MACs: artificial chromosomes as research tools. **Trends in Biotechnology**, v. 12, n. 7, p. 280–6, 1994.

MURRAY, A. W.; SZOSTAK, J. W. Construction of artificial chromosomes in yeast. **Nature**, v. 305, n. 5931, p. 189–93, 1983.

NÄÄTSAARI, L.; MISTLBERGER, B.; RUTH, C.; HAJEK, T.; HARTNER, F. S.; GLIEDER, A. Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. **PLOS One**, v. 7, n. 6, 2012.

NAESBY, M.; NIELSEN, S. V. V.; NIELSEN, C. A. F. *et al.* Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. **Microbial Cell Factories**, v. 8, n. 45, 2009.

NAGAKI, K.; SHIBATA, F.; KANATANI, A.; KASHIHARA, K.; MURATA, M. Isolation of centromeric-tandem repetitive DNA sequences by chromatin affinity purification using a HaloTag7-fused centromere-specific histone H3 in tobacco. **Plant Cell Reports**, v. 31, p. 771–9, 2012.

NAKAMURA, Y.; NISHI, T.; NOGUCHI, R.; ITO, Y.; WATANABE, T.; NISHIYAMA, T.; AIKAWA, S.; HASUNUMA, T.; ISHII, J.; OKUBO, Y.; KONDO, A. A stable, autonomously replicating plasmid vector containing *Pichia pastoris* centromeric DNA. **Applied and Environmental Microbiology**, v. 84, n. 15, 2018.

NETT, J. H.; GERNGROSS, T. U. Cloning and disruption of the *PpURA5* gene

and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*. **Yeast**, v. 20, n. 15, p. 1279–90, 2003.

NETT, J. H.; 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**, v. 22, n. 4, p. 295–304, 2005.

NIELSEN, A. A. K.; SEGALL-SHAPIRO, T. H.; VOIGT, C. A. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. **Current Opinion in Chemical Biology**, v. 17, n. 6, p. 878–92, 2013.

NORDÉN, K.; AGEMARK, M.; DANIELSON, J. A. H.; ALEXANDERSSON, E.; JOHANSON, U. Increasing gene dosage greatly enhances recombinant expression of aquaporins in *Pichia pastoris*. **BMC Biotechnology**, v. 11, n. 47, 2011.

OBST, U.; LU, T. K.; SIEBER, V. A Modular toolkit for generating *Pichia pastoris* secretion libraries. **ACS Synthetic Biology**, v. 6, n. 6, p. 1016-25, 2017.

OHI, H.; OKAZAKI, N.; UNO, S.; MIURA, M.; HIRAMATSU, R. Chromosomal DNA patterns and gene stability of *Pichia pastoris*. **Yeast**, v. 14, n. 10, p. 895-903, 1998.

PANZERI, L.; PHILIPPSSEN, P. Centromeric DNA from chromosome VI in *Saccharomyces cerevisiae* strains. **The EMBO Journal**, v. 1, n. 12, p. 1605-11, 1982.

PEÑA, D. A.; GASSER, B.; ZANGHELLINI, J.; STIEGER, M. G.; MATTANOVISH, D. Metabolic engineering of *Pichia pastoris*. **Metabolic Engineering**, v. 50, p. 2–15, 2018.

PENNANEACH, V.; KOLODNER, R. D. Stabilization of dicentric translocations through secondary rearrangements mediated by multiple mechanisms in *S. cerevisiae*. **PLOS ONE**, v. 4, n. 7, 2009.

PICATAGGIO, S. Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals. **Current Opinion in Biotechnology**, v. 20, n. 3, p. 325-9, 2009.

PINHEIRO, V. B.; TAYLOR, A. I.; COZENS, C.; ABRAMOV, M.; RENDERS, M.;

ZHANG, S.; CHAPUT, J. C.; WENGEL, J.; PEAK-CHEW, S.; MCLAUGHLIN, S. H.; HERDEWIJN, P.; HOLLIGER, P. Synthetic Genetic Polymers Capable of Heredity and Evolution. **Science Reports**, v. 336, p. 341–4, 2012.

PIVA, L. C.; DE MARCO, J. L.; DE MORAES, L. M. P.; REIS, V. C. B.; TORRES, F. A. G. Acetamidase as a dominant recyclable marker for *Komagataella phaffii* strain engineering. **Applied Microbiology and Biotechnology**, v. 102, n. 6, p. 2753–61, 2018.

POTVIN, G.; AHMAD, A.; ZHANG, Z. Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: A review. **Biochemical Engineering Journal**, v. 64, p. 91–105, 2012.

RANDO, O. J.; WINSTON, F. Chromatin and transcription in yeast. **Genetics**, v. 190, n. 2, p. 351–87, 2012.

RÉBORA, K.; LALOO, B.; DAIGNAN-FORNIER, B. Revisiting purine-histidine cross-pathway regulation in *Saccharomyces cerevisiae*: a central role for a small molecule. **Genetics**, v. 170, n. 1, p. 61–70, 2005.

RECH, E. L.; DOBSON, M. J.; DAVEY, M. R.; MULLIGAN, B. J. Introduction of a yeast artificial chromosome vector into *Saccharomyces cerevisiae* cells by electroporation. **Nucleic Acids Research**, v. 18, n. 5, p. 1313, 1990.

REDDEN, H.; MORSE, N.; ALPER, H. S. The synthetic biology toolbox for tuning gene expression in yeast. **FEMS Yeast Research**, v. 15, n. 1, 2014.

REIS, V. C. B.; NICOLA, A. M.; OLIVEIRA NETO, O. de S.; BATISTA, V. D. F.; MORAES, L. M. P. de.; TORRES, F. A. G. Genetic characterization and construction of an auxotrophic strain of *Saccharomyces cerevisiae* JP1, a Brazilian industrial yeast strain for bioethanol production. **Journal of Industrial Microbiology & Biotechnology**, v. 39, n. 11, p. 1673–83, 2012.

RESNICK, M. A.; WESTMORELAND, J.; BLOOM, K. Heterogeneity and maintenance of centromere plasmid copy number in *Saccharomyces cerevisiae*. **Chromosoma**, v. 99, n. 4, p. 281–8, 1990.

RICHARDSON, S. M. *et al.* Design of a synthetic yeast genome. **Science**, n. 355, p. 1040-4, 2017.

RIETHMAN, H. C.; MOYZIS, R. K.; MEYNE, J.; BURKE, D. T.; OLSON, M. V.

Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. **Proceedings of the National Academy of Sciences**, v. 86, n. 16, p. 6240–4, 1989.

RO, D.; PARADISE, E. M.; OUELLET, M.; *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. **Nature**, v. 440, n. 7086, p. 940-3, 2006.

SAMBROOK, J.; FRITSCH, E. F.; MANIATIS, T. **Molecular Cloning: A Laboratory Manual**. 2nd ed. Cold Spring Harbor Laboratory Press, New York, 2001.

SATO, H.; MASUDA, F.; TAKAYAMA, Y.; TAKAHASHI, K.; SAITOH, S. Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. **Current Biology**, v. 22, n. 8, p. 658–67, 2012.

SAUER, B. Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. **Molecular and Cellular Biology**, v. 7, n. 6, p. 2087–96, 1987.

SCHWARZHANS, J. P.; LUTTERMANN, T.; WIBBERG, D.; WINKLER, A.; HUBNER, W.; HUSER, T.; KALINOWSKI, J.; FRIEHS, K. A mitochondrial autonomously replicating sequence from *Pichia pastoris* for uniform high level recombinant protein production. **Frontiers in Microbiology**, v. 8, 2017a.

SCHWARZHANS, J. P.; LUTTERMANN, T.; GEIER, M.; KALINOWSKI, J.; FRIEHS, K. Towards system metabolic engineering in *Pichia pastoris*. **Biotechnology Advances**, v. 35, n. 6, p. 681-710, 2017b.

SCHWARZHANS, 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. **Microbial Cell Factories**, n. 15, v. 84, 2016.

SHAO, Y. *et al.* Creating a functional single-chromosome yeast. **Nature**, v. 560, n. 7718, p. 331-5, 2018.

SHEN, Y. *et al.* SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. **Genome Research**, v. 26, n. 1, p. 36–49,

2016.

SLEISTER, H. M.; MILLS, K. A.; BLACKWELL, S. E.; KILLARY, A. M.; MURRAY, J. C.; MALONE, R. E. Construction of a human chromosome 4 YAC pool and analysis of artificial chromosome stability. **Nucleic acids research**, v. 20, n. 13, p. 3419–25, 1992.

SOLÀ, A.; MAAHEIMO, H.; YLONEN, K.; FERRER, P.; SZYPERSKI, T. Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. **European Journal of Biochemistry**, v. 271, n. 12, p. 2462–70, 2004.

SOLIS-ESCALANTE, D.; KUJIPERS, N. G. A.; BONGAERTS, N.; BOLAT, I.; BOSMAN, L.; PRONK, J. T.; DARAN, J.; DARAN-LAPUJADE, P. *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. **FEMS Yeast Research**, v. 13, p. 126-39, 2013.

STEENSELS, J.; SNOEK, T.; MEERSMAN, E.; PICCA NICOLINO, M.; VOORDECKERS, K.; VERSTREPEN, K. J.. Improving industrial yeast strains: exploiting natural and artificial diversity. **FEMS Microbiology Reviews**, v. 38, n. 15, p. 947–95, 2014.

STIMPSON, K. M.; MATHENY, J. E.; SULLIVAN, B. A. Dicentric chromosomes: unique models to study centromere function and inactivation. **Chromosome Research**, v. 20, n. 5, p. 595–605, 2012.

STOTZ, A.; LINDER, P. The *ADE2* gene from *Saccharomyces cerevisiae*: sequence and new vectors. **Gene**, v. 95, n. 1, p. 91–8, 1990.

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.; GLIEDER, A. Refined *Pichia pastoris* reference genome sequence. **Journal of Biotechnology**, v. 235, p. 121–31, 2016.

SZOSTAK, J.; BLACKBURN, E. H. Cloning Yeast Telomeres on Linear Plasmid Vectors. **Cell**, v. 29, p. 245-55, 1982.

THRASH-BINGHAM, C.; FANGMAN, W. L. A yeast mutation that stabilizes a plasmid bearing a mutated ARS1 element. **Molecular and Cellular Biology**, v.

9, n. 2, p. 809–16, 1989.

TIAN, J.; MA, K.; SAAEM, I. Advancing high-throughput gene synthesis technology. **Molecular BioSystems**, v. 5, p. 714-22, 2009.

TIBBETTS, A S.; APPLING, D. R. Characterization of two 5-aminoimidazole-4-carboxamide ribonucleotide transformylase/inosine monophosphate cyclohydrolase isozymes from *Saccharomyces cerevisiae*. **The Journal of Biological Chemistry**, v. 275, n. 27, p. 20920–7, 2000.

UGOLINI, S.; BRUSCHI, C. V. The red/white colony color assay in the yeast *Saccharomyces cerevisiae*: epistatic growth advantage of white *ade8-18, ade2* cells over red *ade2* cells. **Current Genetics**, v. 30, n. 6, p. 485–92, 1996.

VALLI, M.; TATTO, N. E.; PAYMANN, A.; GRUBER, C.; LANDES, N.; EKKER, H.; THALLINGER, G. G.; MATTANOVISH, D.; GASSER, B.; GRAF, A. B. Curation of the genome annotation of *Pichia pastoris* (*Komagataella phaffii*) CBS7435 from gene level to protein function. **FEMS Yeast Research**, v. 16, n. 6, 2016.

VAROQUAUX, N.; LIACHKO, I.; AY, F.; BURTON, J. N.; SHENDURE, J.; DUNHAM, M. J.; VERT, J.; NOBLE, W. S. Accurate identification of centromere locations in yeast genomes using Hi-C. **Nucleic Acids Research**, v. 43, n. 11, p. 5331–9, 2015.

VARSHNEY, R. K.; RIBAUT, J.; BUCKLER, E. S.; TUBEROSA, R.; RAFALSKI, J. A.; LANGRIDGE, P. A prototrophic deletion mutant collection for yeast metabolomics and systems biology. **Nature Biotechnology**, v. 30, n. 12, p. 1176-8, 2012.

VOGL, T.; GLIEDER, A. Regulation of *Pichia pastoris* promoters and its consequences for protein production. **New Biotechnology**, v. 30, n. 4, p. 385–404, 2013.

VOGL, T.; HARTNER, F. S.; GLIEDER, A. New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. **Current Opinion in Biotechnology**, v. 24, 2013.

WANG, S.; YAO, Q.; TAO, J.; QIAO, Y.; ZHANG, Z. Co-ordinate expression of glycine betaine synthesis genes linked by the FMDV 2A region in a single open

reading frame in *Pichia pastoris*. **Applied Microbiology and Biotechnology**, v. 77, n. 4, p. 891–9, 2007.

WELLINGER, R. ZAKIAN, V. A. Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. **Genetics**, v. 191, p. 1073-1105, 2012.

WENINGER, A.; HATZL, A.; SCHMID, C.; VOGL, T.; GLIEDER, A. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. **Journal of Biotechnology**, DOI: 10.1016/j.jbiotec.2016.03.027, 2016.

WILLIAMSON, D. H. The Yeast ARS Element , Six Years On: A Progress Report. **Yeast**, v. 1, p. 1–14, 1985.

WU, S.; LETCHWORTH, G. J. High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. **BioTechniques**, v. 36, n. 1, p. 152–4, 2004.

XU, L.; WANG, G.; WANG, W.; LY, X.; LI, Q.; YAN, A. Development of a prokaryotic-like polycistronic expression system based on a virus-originated internal ribosome entry site (IRES) in industrial eukaryotic microorganisms. **RSC Advances**, v. 4, n. 93, p. 51615–18, 2014.

YAMADA, Y.; MATSUDA, M.; MAEDA, K.; MIKATA, K. The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of *Komagataella* gen. nov. (*Saccharomycetaceae*). **Bioscience, Biotechnology, and Biochemistry**, v. 59, n. 3, p. 439–44, 1995.

YANG, J.; NIE, L.; CHEN, B.; LIU, Y.; KONG, Y.; WANG, H.; DIAO, L. Hygromycin-resistance vectors for gene expression in *Pichia pastoris*. **Yeast**, v. 31, n. 4, p. 115–25, 2014.

YU, X.; WANG, R.; ZHANG, M.; XU, Y.; XIAO, R. Enhanced thermostability of a *Rhizopus chinensis* lipase by *in vivo* recombination in *Pichia pastoris*. p. 1–11, 2012.

ZHANG, T.; GE, C.; DENG, L.; TAN, T.; WANG, F. C4-dicarboxylic acid production by overexpressing the reductive TCA pathway. **FEMS Microbiology Letters**, v. 362,n. 9, 2015.

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. **Biotechnology Letters**, v. 31, n. 5, p. 679–84, 2009.

8. Anexos

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```

M13-F      -----TTGGT-----
original   -----GTA AACGACGGCCAGTGAGCGCGC
M13-R      ATGCTGATGATCATGATGATATGATGAAGCTGCATGCTTCATCATGATCGATGATGATGA
                                         *

M13-F      -----CTATAGGGCG-ATTGGGTACCAGCGCTGGATGCTGGATGCTGGAT
original   GTAATA--CGACTCACTATAGGGCGAATTGGGTACCAGCGCTGGATGCTGGATGCTGGAT
M13-R      ATGATGCTCGATGATTCATGATGCATGATGCATGCTCCCGCTGGATGATGGATGCTGGAT
                    **  **  **  *  *  *****
M13-F      GCTGGATGCTGGATGCTGG---TGGATGCTG---GATGCTGGATGCTGGATGCTGGATG
original   GCTGGATGCTGGATGCTGGATGCTGGATGCTG---GATGCTGGATGCTGGATGCTGGATG
M13-R      GCTGGATGCTGGATGCTGGATGATGGATGCTGCATGATGCAGGATGCAGGATGCAGGATG
                    *****

M13-F      CTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGG
original   CTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGG
M13-R      CTTGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGG
** *****

M13-F      ATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGC
original   ATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGC
M13-R      ATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGC
                    *****

M13-F      TGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTG---G---T
original   TGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTG---GATGCT
M13-R      TGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCT
                    *****

M13-F      GGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGAT
original   GGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGAT
M13-R      GGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGAT
                    *****

M13-F      GCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTG
original   GCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTG
M13-R      GCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTG
                    *****

M13-F      GATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATG
original   GATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATG
M13-R      GATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATG
                    *****

M13-F      CTGGATGCTGGATGCTGGAGGCTGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCT
original   CTGGATGCTGGATGCTGGAGGCTGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCT
M13-R      CTGGATGCTGGATGCTGGAGGCTGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCTGGATGCT
                    *****

```


CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```

5zeo
original
qAMS-F
-----
GGACTCCTTACAAGCATGATTTTGGTCACGACCTGATTTCCACATCTACGCCGCCGACG
-----TTTCCACATCTACGCCGCCGACG

5zeo
original
qAMS-F
-----
GTTCCGCAGATGTTATGCGTGACATTTCCGCATCTGGAGAGCCTGCCATTCCAACATCA
GTTCCGCAGATGTTATGCGTGACATTTCCGCATCTGGAGAGCCTGCCATTCCAACATCA

5zeo
original
qAMS-F
-----
AAGACTTGTGAATCCTAATATTAAGGCTGTGAATATGAACGAGTTGTGGGATACACATC
AAGACTTGTGAATCCTAATATTAAGGCTGTGAATATGAACGAGTTGTGGGATACACATC

5zeo
original
qAMS-F
-----TACCAAATGGAGTACTTGGAGAAGTGGCGTGAAGCTGAAGAGAAGG
TTCAGAAGTGAATTACCAAATGGAGTACTTGGAGAAGTGGCGTGAAGCTGAAGAGAAGG
TTCAGAAGTGAATTACCAAATGGAGTACTTGGAGAAGTGGCGTGAAGCTGAAGAGAAGG
*****

5zeo
original
qAMS-F
CTGGAAGGAACCTTGACGCCATCATCGCCCTATTACTCCAACCTGCCGCTGTTTCGTCATG
CTGGAAGGAACCTTGACGCCATCATCGCCCTATTACTCCAACCTGCCGCTGTTTCGTCATG
CTGGAAGGAACCTTGACGCCATCATCGCCCTATTACTCCAACCTGCCGCTGTTTCGTCATG
*****

5zeo
original
qAMS-F
ATCAATTCGGTACTACGGTTACGCTTCCGTTATCAACTTGTGATTTCACATCCGTTG
ATCAATTCGGTACTACGGTTACGCTTCCGTTATCAACTTGTGATTTCACATCCGTTG
ATCAATTCGGTACTACGGTTACGCTTCCGTTATCAACTTGTGATTTCACATCCGTTG
*****

5zeo
original
qAMS-F
TTGTCCCTGTGACATTCGCTGATAAGAACATTGACAAGAAGAATGAATCTTCAAGGCCG
TTGTCCCTGTGACATTCGCTGATAAGAACATTGACAAGAAGAATGAATCTTCAAGGCCG
TTGTCCCTGTGACATTCGCTGATAAGAACATTGACAAGAAGAATGAATCTTCAAGGCCG
*****

5zeo
original
qAMS-F
TCTCCGAACCTGGACGCTTTGGTGCAAGAAGAGTACGATCTTGAAGCCTACCACGGTGC TC
TCTCCGAACCTGGACGCTTTGGTGCAAGAAGAGTACGATCTTGAAGCCTACCACGGTGC TC
TCTCCGAACCTGGACGCTTTGGTGCAAGAAGAGTACGATCTTGAAGCCTACCACGGTGC TC
*****

5zeo
original
qAMS-F
CAGTCGCTGTCCAAGTCATCGGTCGTCGTTTGTCCGAAGAGCGTACCTTGGCCATTGCTG
CAGTCGCTGTCCAAGTCATCGGTCGTCGTTTGTCCGAAGAGCGTACCTTGGCCATTGCTG
CAGTCGCTGTCCAAGTCATCGGTCGTCGTTTGTCCGAAGAGCGTACCTTGGCCATTGCTG
*****

5zeo
original
qAMS-F
AAGAAGTTGGTAACTGTTGGGTAATGTCGTGACTCCATAGCACGT-CCGACGGCGGCC
AAGAAGTTGGTAACTGTTGGGTAATGTCGTGACTCCATAGCACGT-CCGACGGCGGCC
AAGAAGTTGGTAACTGTTGGGTAATGTCGTGACTCCATAGCACGTCCCGACGGCGGCC
*****

```

5zeo original qAMS-F	ACGGGTCTCAGACGTCTGGGATCCGTC----- ACGGGTCCCAGGCCTCGGAGATCCGTC----- ACGGGTCCCAGGCCTCGGAGATCCGTC----- ***** ** * ** * *****
5zeo original qAMS-F	----- TATGTCACGCTTACATTACGCCCTCCCCCACATCCGCTCTAACGAAAAGGAAGGAGT -----
5zeo original qAMS-F	----- TAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTATAGTTATGTTAGTATTAAGAAC -----
5zeo original qAMS-F	----- GTTATTTATATTTCAAATTTTTCTTTTTTTCTGTACAGACGCGTGTACGCATGTAACAT -----
5zeo original qAMS-F	----- TATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCCATAT -----
5zeo original qAMS-F	----- GGCGGCCGCTTGCTCACATGTTGGTCTCCAGCTTGCAAATTAAGCCTTCGAGCGTCCCA -----
5zeo original qAMS-F	----- AAACCTTCTCAAGCAAGGTTTTCAGTATAATGTTACATGCGTACACGCGTCTGTACAGAA -----
5zeo original qAMS-F	----- AAAAAAGAAAAATTTGAAATATAAATAACGTTCTTAATACTAACATAACTATAAAAAAAT -----
5zeo original qAMS-F	----- AAATAGGGACCTAGACTTCAGGTTGTCTAACTCCTTCTTTTCGGTTAGAGCGGATGTGG -----
5zeo original qAMS-F	----- GGGGAGGGCGTGAATGTAAGCGTGACATAACTAATTACATGATATCGACAAAGGAAAAGG -----

```

5zeo          -----GGGCCG-CGTCGGACGTGTCAGTCCTGCTCC
original      GGGACGGATCTCCGAGGCTGGGACCCGTGGGCCGCCGTGGACGTGTCAGTCCTGCTCC
qAMS-F        -----GGGCCG-CGTCGGACGTGTCAGTCCTGCTCC
                *****

5zeo          TCGGCCACGAAGTGCACGCAGTTGCCGGCCGGGTGCGCGAGGGCGAACTCCCGCCCCAC
original      TCGGCCACGAAGTGCACGCAGTTGCCGGCCGGGTGCGCGAGGGCGAACTCCCGCCCCAC
qAMS-F        TCGGCCACGAAGTGCACGCAGTTGCCGGCCGGGTGCGCGAGGGCGAACTCCCGCCCCAC
                *****

5zeo          GGCTGCTCGCCGATCTCGGTCATGGCCGGCCGGAGGCGTCCCGGAAGTTCGTGGACACG
original      GGCTGCTCGCCGATCTCGGTCATGGCCGGCCGGAGGCGTCCCGGAAGTTCGTGGACACG
qAMS-F        GGCTGCTCGCCGATCTCGGTCATGGCCGGCCGGAGGCGTCCCGGAAGTTCGTGGACACG
                *****

5zeo          ACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTG
original      ACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTG
qAMS-F        ACCTCCGACCACTCGGCGTACAGCTCGTCCAGGCCGCGCACCCACACCCAGGCCAGGGTG
                *****

5zeo          TTGTCCGGCACCACCTGGTCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACC
original      TTGTCCGGCACCACCTGGTCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACC
qAMS-F        TTGTCCGGCACCACCTGGTCTGGACCGCGCTGATGAACAGGGTCACGTCGTCCCGGACC
                *****

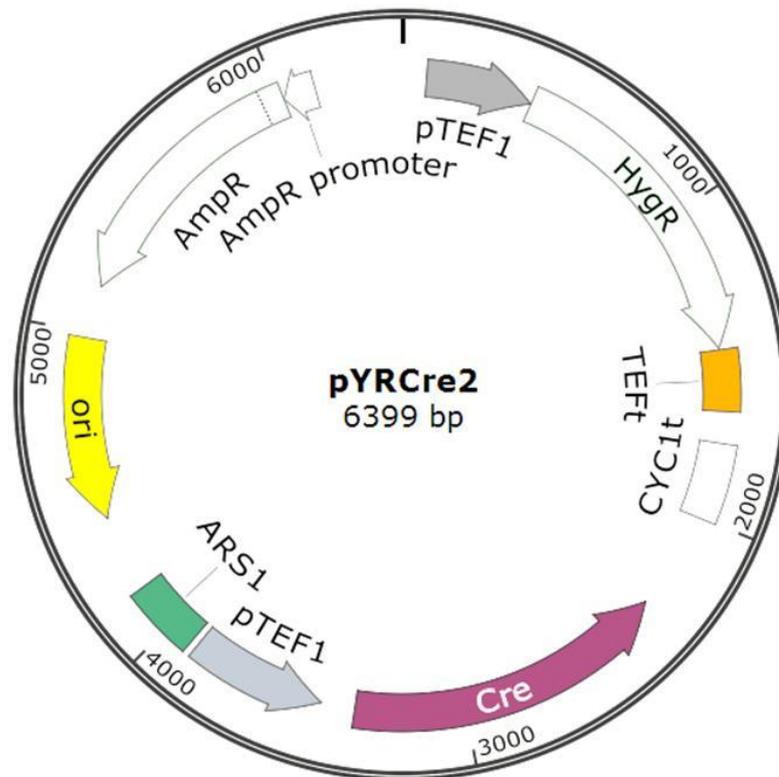
5zeo          ACACCGGCGAAGTCGTCTCCACGAAGACCCGGGAGAACCAGCCGGTCCGAGTCCAGAAC
original      ACACCGGCGAAGTCGTCTCCACGAAGTCCCGGGAGAACCAGCCGGTCCGAGTCCAGAAC
qAMS-F        ACACCG-----
                *****

5zeo          TCGACCGCTCCGGCGACGTCGCGC-----
original      TCGACCGCTCCGGCGACGTCGCGCGCGGTGAGCACCGGAACGG
qAMS-F        -----

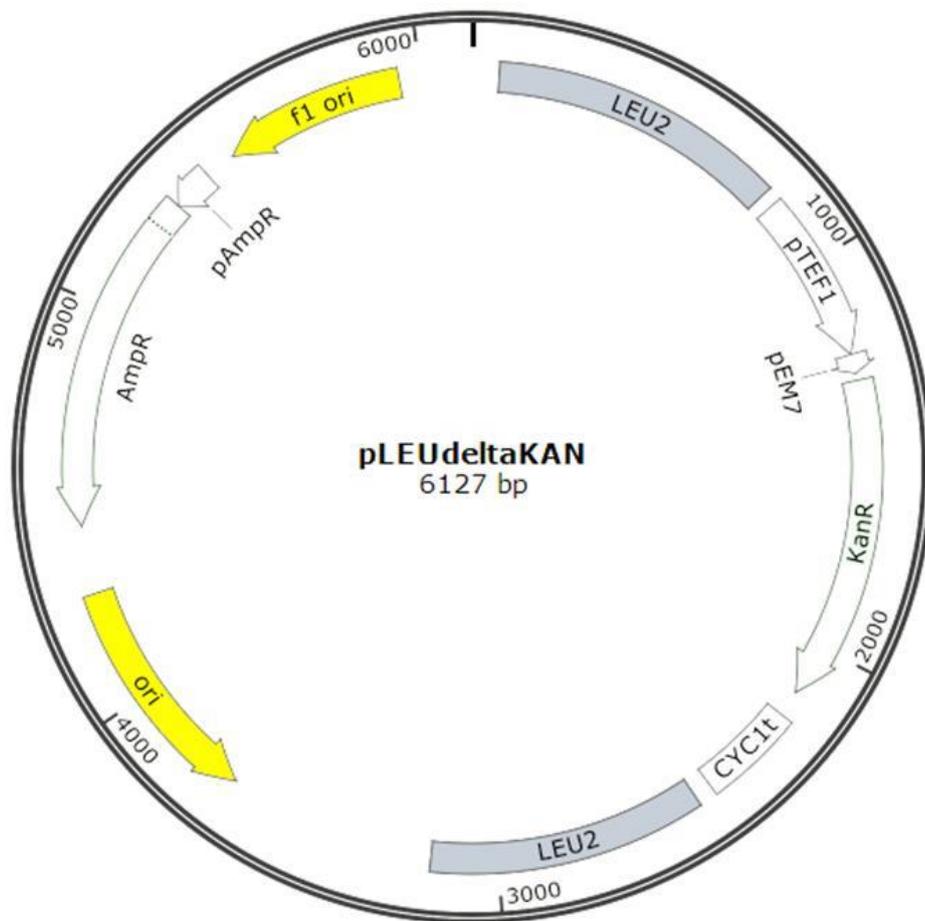
```

Anexo 2. Alinhamento das sequências obtidas pelo sequenciamento com a montagem teórica da região da PCR de confirmação número 7.

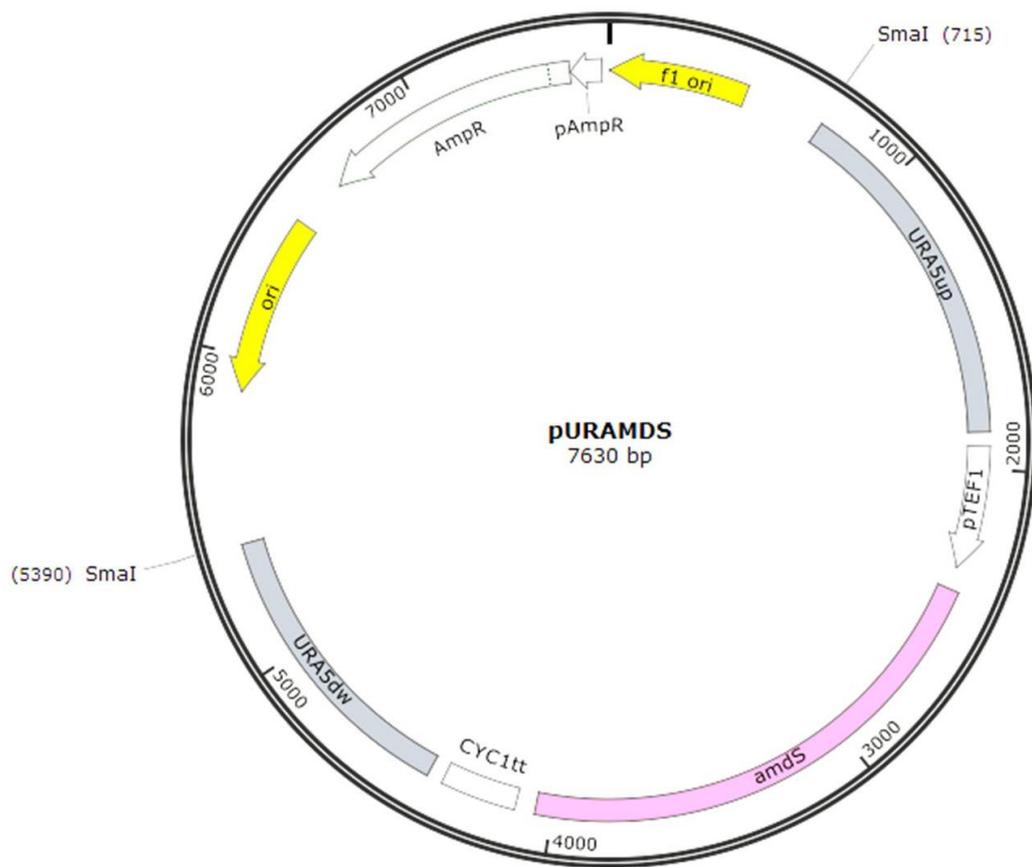
Sequenciamento foi realizado utilizando os *primers* 5zeo e qAMS-F e comparado com a sequência original (ao centro). Sequências realçadas em amarelo, ausentes nos sequenciamentos, correspondem aos dois terminadores CYC1 dos cassetes de expressão dos genes *Sh ble* e *amdS*.



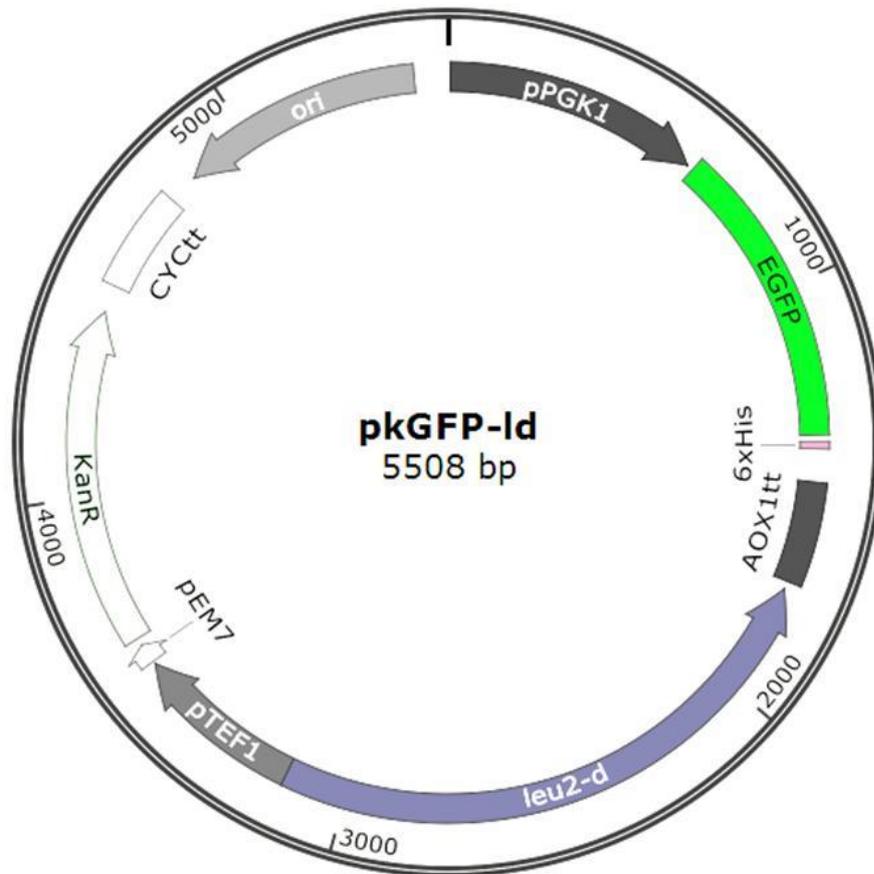
Anexo 3. Mapa do **vetor pYRCre2**. Em roxo, o gene da CreA recombinase sob controle do promotor *TEF1* (cinza) e terminador *CYC1* (branco). Em verde, sequência ARS de *S. cerevisiae*. Genes de seleção HygR que confere resistência a higromicina B e AmpR de resistência a ampicilina estão indicados em branco. Em amarelo, a origem de replicação de *E. coli*. Baseado em BETANCUR *et al.*, (2017).



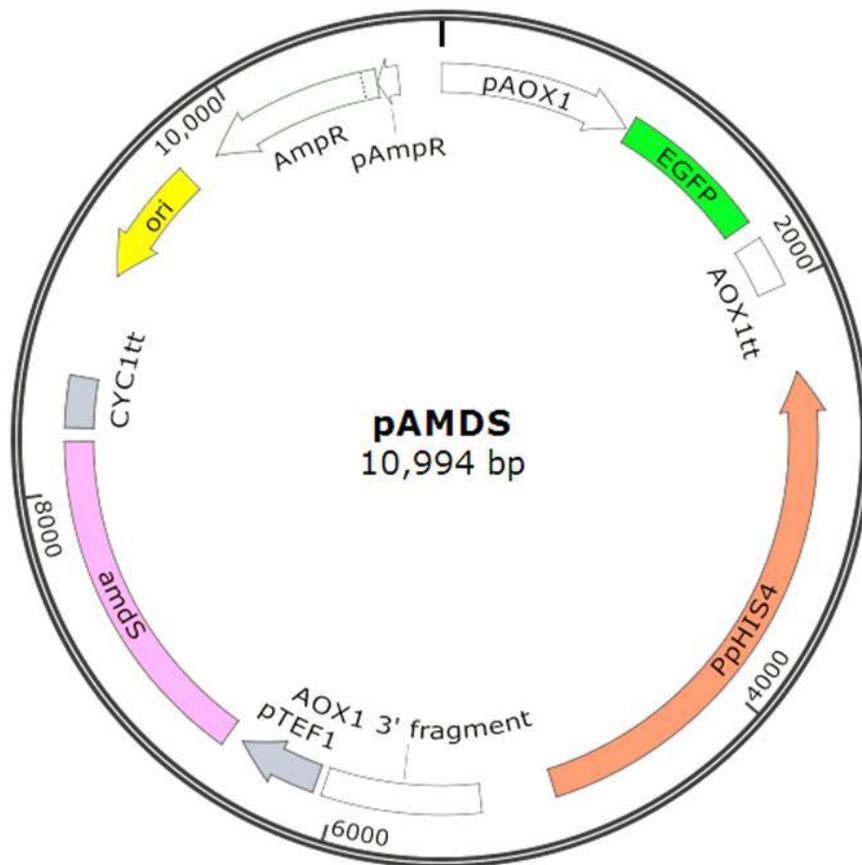
Anexo 4. Mapa do vetor pLEUdeltaKAN. Em branco, o cassete de resistência a G418 (*kanR*). Em cinza, as sequências utilizadas para orientar a integração no *locus LEU2*. Sequências de *E. coli* estão indicadas em amarelo e o cassete de resistência a ampicilina (AmpR), em branco. Baseado em BETANCUR *et al.*, (2017)



Anexo 5. Mapa do vetor pURAMDS. Em rosa está o gene *amdS* de *A. nidulans* sob controle do promotor TEF1 e terminador CYC1 (em branco) de *S. cerevisiae*. Em cinza são indicadas as sequências usadas para orientação da integração no *locus URA5*. Baseado em PIVA *et al.*, 2018.



Anexo 6. Mapa do vetor pkEGFP-Id. Gene EGFP é indicado em verde, sob o controle do promotor *PGK1* e terminador *AOX1* de *K. phaffii*. Em branco, a marca de resistência a G418 (*KanR*) sob controle do promotor *TEF1* e terminador *CYC1* de *S. cerevisiae*. Baseado em BETANCUR *et al.*, 2017.



Anexo 7. Mapa do plasmídeo pAMDS. Gene *amdS* de *A. nidulans* é indicado em rosa, sob o controle do promotor *TEF1* e terminador *CYC1* de *S. cerevisiae* (em cinza). Marca auxotrófica *HIS4* de *K. phaffii* é indicada em alaranjado, gene *EGFP* está em verde sob controle do promotor e terminador *AOX1* em branco. Baseado em PIVA *et al.*, 2018.



Acetamidase as a dominant recyclable marker for *Komagataella phaffii* strain engineering

Luiza Cesca Piva¹ · Janice Lisboa De Marco¹ · Lidia Maria Pepe de Moraes¹ · Viviane Castelo Branco Reis¹ · Fernando Araripe Gonçalves Torres¹

Received: 10 November 2017 / Revised: 18 January 2018 / Accepted: 20 January 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

We have investigated the use of the gene coding for acetamidase (*amdS*) as a recyclable dominant marker for the methylotrophic yeast *Komagataella phaffii* in order to broaden its genetic toolbox. First, the endogenous constitutive *AMD2* gene (a putative acetamidase) was deleted generating strain LA1. A cassette (*amdSloxP*) was constructed bearing a codon-optimized version of the *Aspergillus nidulans amdS* gene flanked by *loxP* sites for marker excision with Cre recombinase. This cassette was successfully tested as a dominant selection marker for transformation of the LA1 strain after selection on plates containing acetamide as a sole nitrogen source. Finally, *amdSloxP* was used to sequentially disrupt the *K. phaffii ADE2* and *URA5* genes. After each disruption event, a Cre-mediated marker recycling step was performed by plating cells on medium containing fluoroacetamide. In conclusion, *amdS* proved to be a suitable tool for *K. phaffii* transformation and marker recycling thus providing a new antibiotic-free system for genetic manipulation of this yeast.

Keywords Acetamidase · Marker recycling · *amdS* · Dominant marker · *Komagataella phaffii*

Introduction

Komagataella phaffii (formerly *Pichia pastoris*) is a methylotrophic yeast which has important biotechnological applications, mainly as a host for heterologous gene expression for the production of a wide range of microbial, plant, and human proteins (Cereghino and Cregg 2000; Cereghino et al. 2002; Potvin et al. 2012). Genetic transformation of *K. phaffii* is typically based on integrative plasmids carrying either auxotrophic or dominant markers (Celik and Calik 2012). Auxotrophic markers which have been used in *K. phaffii* include genes involved in nucleotide synthesis (*URA3*, *URA5*, *ADE1*, and *ADE2*), amino acid production (*ARG1*, *ARG2*, *ARG3*, *MET2*, *HIS1*, *HIS2*, *HIS4*, *HIS5*, *HIS6*, and *LEU2*) and other biosynthetic pathways (*FLD1*) (Betancur et al.

2017; Lin Cereghino et al. 2001; Nett and Gerngross 2003; Nett et al. 2005; Sunga and Cregg 2004; Thor et al. 2005). Unlike auxotrophic markers which require specific strains for their use, dominant markers may be employed more broadly. In *K. phaffii*, these markers are typically represented by drug-resistance genes such as *hph* (hygromycin B), *Sh ble* (zeocin), *BSD* (blasticidin S), *kan* (G418) and (Delic et al. 2010; Drocourt et al. 1990; Kimura et al. 1994; Scorer et al. 1994). Also, the *Saccharomyces cerevisiae SUC2* (invertase) gene has been used as a dominant marker in *K. phaffii* allowing growth in medium containing sucrose as sole carbon source (Sreekrishna et al. 1987).

Some markers may also be used in counter-selection approaches to eliminate unwanted DNA sequences when transformed cells are exposed to restrictive conditions. For example, the drug 5-fluoroorotic acid (5-FOA) is commonly used to select cells that lose either *URA3* or *URA5* (Boeke et al. 1984). This is a useful genetic strategy for marker recycling when different transformation events have to be performed in the same strain and the availability of transformation vectors is limited. In *K. phaffii*, *URA5* and the double markers *ARG4/TurJ3* and *mazF-ZeoR* have been successfully used in counter-selection approaches (Nett and Gerngross 2003; Soderholm et al. 2001; Yang et al. 2009). Removal of the selection marker

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00253-018-8824-9>) contains supplementary material, which is available to authorized users.

✉ Fernando Araripe Gonçalves Torres
ftorres@unb.br

¹ Instituto de Ciências Biológicas, Departamento de Biologia Celular, Universidade de Brasília, Brasília, DF 70910-900, Brazil

gene in marker recycling strategies has been performed in *P. pastoris* using Cre recombinase, which removes DNA marked with *loxP* sequences (Marx et al. 2008).

The *amdS* gene which codes for acetamidase (EC 3.5.1.4) in *Aspergillus nidulans* has been used as a dominant selection marker in many species of filamentous fungi as well as in some yeasts, such as *Candida glabrata*, *Kluyveromyces lactis*, and *S. cerevisiae* (Fu et al. 2016; Gomi et al. 1992; Jørgensen et al. 2014; Read et al. 2007; Solis-Escalante et al. 2013). Acetamidase catalyzes the hydrolysis of acetamide into acetate and ammonia thus making this last compound available as a nitrogen source. Like *URA3/URA5*, *amdS* also has a counter-selection property—cells that have lost the marker can be selected in the presence of the drug fluoroacetamide, which is hydrolyzed by acetamidase into ammonia and the toxic compound fluoroacetate (Apirion 1965; Hynes 1970).

In this study, we assessed the effectiveness of *amdS* as a dominant and counter-selectable marker in *K. phaffii* aiming at the development of a new genetic tool for metabolic engineering in this yeast.

Materials and methods

Strains and media

K. phaffii X-33 (Invitrogen, Madison, USA) was used as host for yeast transformation. Plasmid cloning and propagation were performed in chemically competent *Escherichia coli* XL10-Gold (Agilent Technologies, Santa Clara, USA). Bacteria were grown in LB medium (5 g L⁻¹ yeast extract, 10 g L⁻¹ peptone, and 10 g L⁻¹ NaCl pH 7.2). Resistant bacteria were selected using 25 µg/mL zeocin or 100 µg/mL ampicillin. Low-salt LB medium (5 g L⁻¹ NaCl instead of 10 g L⁻¹) was used when zeocin was the antibiotic of choice. Yeast strains were grown under non-selective conditions in YPD medium (10 g L⁻¹ yeast extract, 20 g L⁻¹ peptone, and 20 g L⁻¹ glucose). Hygromycin B (200 µg/mL) and zeocin (100 µg/mL) were added to YPD medium when needed. When yeast cells were electroporated with a plasmid carrying a drug resistance marker, 1 M sorbitol was added to YPD medium (YPDS). For transformation selection and RNA extraction, cells were grown on minimal dextrose (MD) medium [34 g L⁻¹ Yeast Nitrogen Base w/o amino acids and ammonium sulfate (Difco, Detroit USA), 20 g L⁻¹ glucose, 100 g L⁻¹ (NH₄)₂SO₄, and 0.0004 g L⁻¹ biotin]. The nitrogen source was replaced by urea (20 g L⁻¹) or acetamide (0.6 g L⁻¹) for RNA extraction. Counter-selection of recycled markerless cells was performed on MD medium containing 100 g L⁻¹ (NH₄)₂SO₄ and 2.3 g L⁻¹ fluoroacetamide. When needed, MD medium was supplemented with 0.2% uracil, 0.002% adenine, or 0.1% 5-FOA. Induction of the *EGFP* gene was performed using minimal methanol (MM) medium (34 g L⁻¹

Yeast Nitrogen Base w/o amino acids and ammonium sulfate, 100 g L⁻¹ (NH₄)₂SO₄, 20 g L⁻¹ methanol, and 0.0004 g L⁻¹ biotin). Solid medium was prepared by the addition of 2% agar to the media. Routinely, cell cultures were grown at 37 °C (bacteria) or 28–30 °C (yeast) with agitation frequency of 200 rpm.

RNA extraction and RT-PCR

Total RNA was extracted with RNeasy Mini Kit (Qiagen, Hilden Germany) or with SV Total RNA Isolation System (Promega, Madison USA). RNA was treated with RQ1 RNase-Free DNase (Promega) following two extractions with phenol/chloroform and ethanol precipitation. cDNA was synthesized by using SuperScript III First-Strand Synthesis System Kit (Invitrogen).

Plasmid construction

All oligonucleotide primers used in this work are shown in Table 1. The *A. nidulans amdS* gene was synthesized by Epoch Life Sciences (Missouri City USA). Introns were removed, and the sequence was codon-optimized for expression in both *K. phaffii* and filamentous fungi (accession no. MG571440). *NcoI* and *StuI* restriction sites were used for subcloning *amdS* into plasmid pGZL (Reis et al. 2012) under the control of the *TEF1* promoter and *CYC1* terminator sequence from *S. cerevisiae*, resulting in plasmid pGamd. The expression cassette was flanked by *loxP* sites for marker excision. The resulting *amdSloxP* cassette was removed from pGamd by digestion with *BamHI* following subcloning into pPIC-B linearized with *BglII*. pPIC-B is based on pPIC9 (Invitrogen) lacking the *BglII* site at position 5621. The resulting plasmid, pAMDS, was digested with *NotI* and *BamHI* for excision of the α -factor secretion signal sequence which was replaced by the enhanced green fluorescent protein gene (*EGFP*) derived from pEGFP-N3 (Clontech, Mountain View, USA) after digestion with the same enzymes. The resulting plasmid was named pAMDS-EGFP.

Two plasmids were used for marker excision using Cre recombinase: pYRCre2 (Betancur et al. 2017) and pPICH-Cre. The former is a replicative vector bearing the hygromycin resistance marker and the Cre recombinase gene under control of the *S. cerevisiae* constitutive *TEF1* promoter and the *CYC1* terminator. The latter was based on pPICHOLI-1 (MoBiTec, Goettingen, Germany). Briefly, primers pARS1 and pPIC-R were used to amplify the region between ARS1 and selection marker sequences on pPICHOLI-1. This reaction removed a *BamHI* site from the ARS1 sequence and added an *AvrII* site. The resulting amplicon was re-amplified with primers pPIC-F and pPIC-R which added *PmeI*, *EcoRI*, *NotI*, and *BamHI* sites. The amplicon was digested with *EcoRI* and re-

Table 1 Primers used in this study. Restriction sites are underlined

Primer	5'-3' sequence	Restriction sites
kpAMS-F4	TTGGTCAATTGTTGTATTGAAATCT	
kpAMS-R4	ACAGAACCTTGACAGCCGATC	
kpAMS-F1	CTGATCGAATTATCAAAATAACGT	
kpAMS-R1	TTCTTTGAGATCTCTACCGAATG	<i>Bgl</i> II
kpAMS-F2	TTT <u>AGATCT</u> GTAAAGTCCATTCAACGCCCAT	<i>Bgl</i> II
kpAMS-R2	CTGAACTCTCTATAACCGACATAT	
kpAMS-F3	TGGATCCGGGAAGGTGATTCAATTTATGTG	
kpAMS-R3	TGGATCCGGGTAATTGAAGATTCAGCACA	
qAMS-F	CTGGACACACCGTCACACCTT	
qAMS-R	GATGTTTGGAAATGGCAGGCTCTC	
ADE2upF	GGGTCTGTAGAAGGTCTATGGA	
ADE2upR	AGAGATCTGTTTCTACGATCTTGGAAAAGGT	<i>Bgl</i> II
ADE2dwF	AGAGATCTGTTCCGGTAGAAATTGTCTCCA	<i>Bgl</i> II
ADE2dwR	GGGACGACGAGAAACATGACGAC	
ADE2F1	TCGTTACGGTCTCTCCAGCTT	
ADE2R1	GGGCACTGAAGACACAGCAAATTC	
URA5upF	<u>TCCCGGG</u> AGGATTGTTGGGTTACGCTATAT	<i>Sma</i> I
URA5upR	AGAGATCTAGCCAGCTCTGATAAAGAACCA	<i>Bgl</i> II
URA5dwF	AGAGATCTGCTCAGGTTCCCTTTCATAAAC	<i>Bgl</i> II
URA5dwR	<u>TCCCGGG</u> CTGAGAGCGCAGTCATCGGA	<i>Sma</i> I
URA5F	CGCTCTTTCCTCTGCTGATGTTT	
URA5R	ATCTTGAGCCTGAGTGCTTCA	
pARS1	AAACCTAGGCAATTAATATTACTTATTTGGTCAA CCCCAAATAGGTTGATTCATACTTGGTTCATTC AAAAATAAGTAGTCTTTTGAGATCTTCAATATT ATAATAATATACTATAACAGCCGACTTGTTC TTTTTCGCGAATGTTCCCGCAGCTTATCACTAGTC CCCACACACCATAGCTTCAAATG	
PPIC-F	GATTTCAGAAAT <u>TCGTTTAAACGCGGCCGCCCTAGGC</u> <u>AATTAATATTTACTTATTTGG</u>	<i>Eco</i> RI <i>Pme</i> I <i>Not</i> I
PPIC-R	GATTTCAGAAAT <u>TCGGATCCGTTTAAACT</u> CATTCAACC TTTCGTCTTTGG	<i>Eco</i> RI <i>Bam</i> HI <i>Pme</i> I

ligated generating plasmid pPICH. The Cre expression cassette was removed from pYRCre2 with *Not*I and subcloned into pPICH digested with the same enzyme resulting in pPICH-Cre (Online Resource 1).

Construction of deletion cassettes

The *AMD2*, *ADE2*, and *URA5* deletion cassettes were constructed using a PCR-based strategy. First, upstream and downstream sequences used for targeted homologous recombination were amplified from the *K. phaffii* genome with Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, USA) using the following primer pairs: kpAMS-F1/kpAMS-R1 and kpAMS-F2/kpAMS-R2 for *AMD2*; ADE2upF/ADE2upR and ADE2dwF/ADE2dwR for

ADE2; and URA5upF/URA5upR and URA5dwF/URA5dwR for *URA5*. Amplicon pairs were digested with *Bgl*II and ligated to each other. The ligation product was amplified by using the external primers for each construction. Amplicons were cloned into the *Sma*I-linearized pBluescript II SK⁺ plasmid. Each plasmid was then digested with *Bgl*II for subcloning of the selected marker cassette isolated from the containing plasmid after digestion with *Bam*HI. For construction of the *AMD2* deletion cassette, we used the *Sh ble* selection marker from plasmid pGZL (Reis et al. 2012). *ADE2* and *URA5* deletion cassettes used the *amdS* marker from plasmid pGamd.

Prior to yeast transformations, *ADE2* and *URA5* deletion cassettes were released from the plasmids after digestion with *Sma*I while the *AMD2* deletion cassette was amplified with primers KpAMS-F1 and KpAMS-R2.

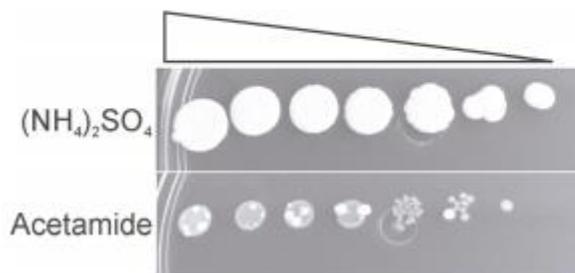


Fig. 1 *K. phaffii* grows on acetamide as a sole N source. Serial dilutions (10^9 to 10^3 cells mL^{-1}) of a *K. phaffii* X-33 culture spotted on MD medium plates supplied with ammonium sulfate (100 g L^{-1}) or acetamide (0.6 g L^{-1}) as sole nitrogen sources. Growth was carried out at 28°C for 7 days

Yeast transformation and marker recycling

Yeast integrative transformation was performed with linearized DNA sequences following the transformation protocol provided in the Pichia Expression Kit (Invitrogen).

For marker recycling, yeast cells were transformed with 100 ng intact pYRCre2 and pPICH-Cre plasmids according to the pPICHOLI instruction manual (MoBiTec). Selected transformed cells were transferred to a fresh plate following colony PCR to confirm marker removal.

GFP detection by fluorescence analysis

Clones were plated on MM medium to induce expression of the *EGFP* gene from the *AOXI* promoter. Fluorescence was detected by scanning plates on Typhoon 9210 Variable Mode Imager (GE Healthcare Life Sciences, Marlborough, USA) using the following parameters: emission filter 526 SP Fluorescein, PMT voltage of 800 V, green laser, and normal sensitivity.

Results

K. phaffii expresses a putative amidase gene

In order to assess the use of *amdS* as a dominant marker in *K. phaffii*, we first investigated acetamide utilization by this yeast as sole nitrogen source. For that purpose, a serial dilution of yeast cells was plated on MD medium containing acetamide or ammonium sulfate (positive control) following incubation at 28°C for 7 days. As expected, full growth was accomplished on ammonium sulfate, whereas sporadic colonies were observed on acetamide (Fig. 1). Considering that *K. phaffii* could be expressing an enzyme with acetamidase activity, genome analysis was carried out to check for this possibility. According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, *K. phaffii* is not capable of producing ammonia from the hydrolysis of acetamide or formamide. Nevertheless, a BLAST search was performed to search for protein sequences that could possibly play this role. The *amdS* sequence from *A. nidulans* (gene ID: 2868318) was used as a query sequence against the *K. phaffii* GS115 genome. The search resulted in the identification of ORF PAS_chr3_0283 (currently *AMD2*), bearing 43% similarity with *amdS*.

In order to investigate whether *AMD2* is transcriptionally active, RT-PCR analysis was performed under different growth conditions. Yeast cells were grown for 2 days in complex YPD medium or MD medium supplied with acetamide, urea, or ammonium sulfate as nitrogen sources for RNA extraction. RT-PCR using primers kpAMS-F4 and kpAMS-R4 resulted in an amplicon of expected ~ 700 bp in complex medium (Fig. 2a) and in MD media with different nitrogen sources (Fig. 2b) indicating that *AMD2* is likely to be constitutively expressed.

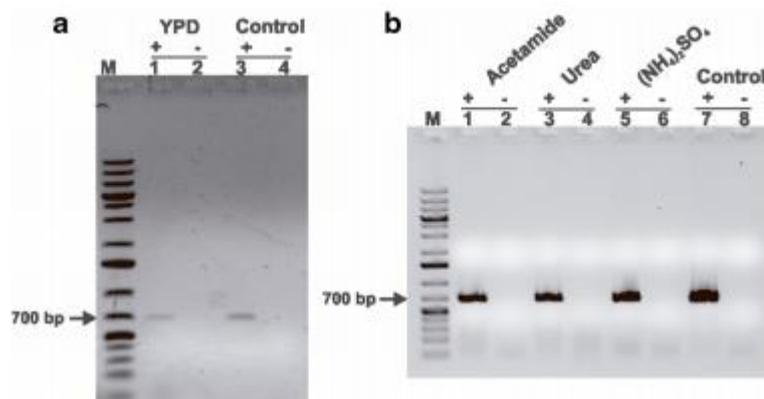
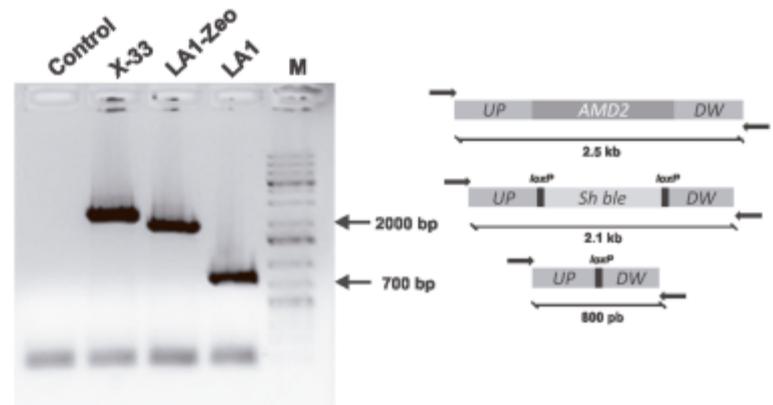


Fig. 2 Transcription profiles of *AMD2* in different media. *K. phaffii* RNA was extracted from yeast cells grown in different media and submitted to RT-PCR following agarose gel electrophoresis analysis. **a** YPD medium, lane 1: (+) RT; lane 2: (-) RT; lane 3: PCR control reaction with genomic DNA; lane 4: PCR control reaction without template. **b** MD medium with

acetamide, urea, or ammonium sulfate as nitrogen sources. Total RNA was submitted (lanes 1, 3, and 5) or not (lanes 2, 4, and 6) to RT prior to PCR. Lanes 7 and 8 represent the PCR control reactions. M: Invitrogen GeneRuler 1 kb Plus

Fig. 3 Deletion of *AMD2* and marker recycling. Gel electrophoresis analysis of colony PCR products using primers KpAMS-F3 and KpAMS-R3 and different DNA templates. Lane 1: no template; lane 2: X-33; lane 3: LA1-Zeo; lane 4: LA1; M: Invitrogen GeneRuler 1 kb Plus. The panel on the right is a schematic representation of expected amplicon sizes



Disruption of *AMD2*

Given that *AMD2* is transcriptionally active and shares similarity to other amidases, we reasoned that it could be related to the sporadic *K. phaffii* colonies observed on the acetamide plate. This phenotypic behavior prompted us to delete *AMD2* in *K. phaffii* X-33 using a cassette bearing the *Sh ble* marker. Clones were isolated in selective medium and analyzed by colony PCR using primers KpAMS-F3 and KpAMS-R3, which anneal externally to the regions used for recombination. A selected clone which yielded the expected 2.1-kb amplification product which corresponded to disrupted *AMD2* was submitted to marker recycling, and the resulting *amd2* strain was named LA1 (Fig. 3). Finally, when LA1 was grown on medium containing acetamide, sporadic colonies were no longer observed (Fig. 4).

Use of *amdS* as a selection marker

To test *amdS* as a dominant marker for *K. phaffii*, LA1 and X-33 strains were transformed with the integrative plasmid pAMDS-EGFP which was targeted to the *HIS4* locus after linearization with *Safl*. In addition to *amdS*, this plasmid also carries the *EGFP* reporter gene to further confirm successful integration. After electroporation, cells were plated on MD medium with acetamide as sole nitrogen source following

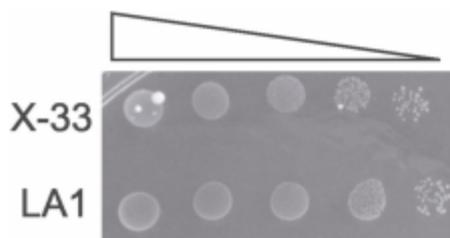


Fig. 4 Acetamide consumption by LA1 strain. Serial dilutions (10^8 to 10^4 cells mL^{-1}) of *K. phaffii* X-33 or LA1 cultures were spotted on MD medium plates supplied with acetamide (0.6 g L^{-1}). Growth was carried out at 28°C for 7 days

incubation at 28°C for 4 days. Control plates containing untransformed LA1 cells showed no growth while X-33 gave rise to 23 colonies. With the pAMDS-EGFP plasmid, both strains yielded roughly the same number of colonies (~ 200). Fifty individual transformed colonies from each system were submitted to fluorescence analysis to confirm the presence of the EGFP protein. Again, both systems provided similar results: 100% positive clones for LA1 and 98% for X-33 transformants. The presence of the heterologous sequences was also confirmed in selected clones by colony PCR using primers qAMS-F and qAMS-R.

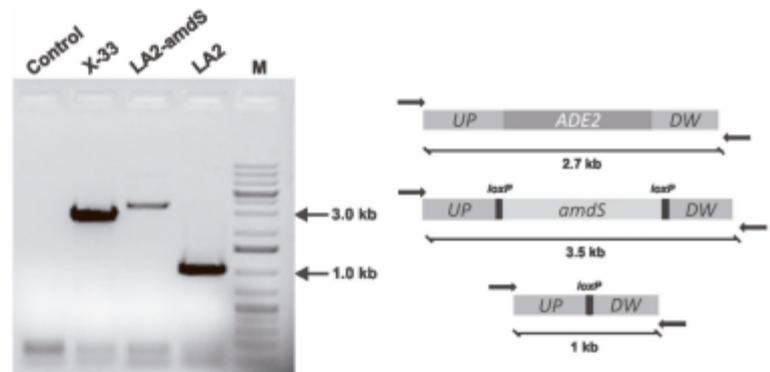
Gene deletion and marker recycling

Since *amdS* was efficiently employed to select *K. phaffii* transformants, we decided to test the *amdSloxP* cassette for gene knockout and marker recycling. Two deletion cassettes were constructed by PCR to sequentially delete the *K. phaffii* *ADE2* and *URA5* genes.

The LA1 strain was first transformed with the *ADE2* deletion cassette. Selection in MD medium with acetamide as nitrogen source resulted in ~ 500 colonies. Three red colonies were screened by colony PCR using confirmation primers Ade2F1 and Ade2R1 which anneal outside the regions used for recombination. Results showed that two of these clones had the *ADE2* gene successfully deleted, providing an amplicon of 3.5 kb instead of the 2.7-kb amplicon expected in the X-33 strain (Fig. 5). One positive clone (LA2-*amdS*) was selected for phenotypic analysis in MD medium which confirmed adenine auxotrophy (data not shown).

In order to remove the *amdS* marker for another gene disruption cycle, the auxotrophic LA2-*amdS* strain was transformed with pYRCre2. Clones selected in YPD containing hygromycin B were transferred to MD medium plates containing fluoroacetamide to test the counter-selection property of the *amdS* marker. Unlike the control LA2-*amdS* strain, all clones could grow in the presence of fluoroacetamide, indicating that the *amdS* marker was no longer present. One of these

Fig. 5 Deletion of *ADE2* and marker recycling. Gel electrophoresis analysis of colony PCR products using primers Ade2F1 and Ade2R1 and different DNA templates. Lane 1: no template; lane 2: X-33; lane 3: LA2-*amdS*; lane 4: LA2; M: Invitrogen GeneRuler 1 kb Plus. The panel on the right is a schematic representation of expected amplicon sizes



clones was also submitted to colony PCR with primers Ade2F1 and Ade2R1 for confirmation of *amdS* marker recycling. This reaction resulted in a 1-kb fragment corresponding to the *ADE2*-disrupted locus without the *amdS* marker (Fig. 5).

After marker recycling, the resulting strain, called LA2, was submitted to a second gene deletion event, this time targeted to the *URA5* locus. The acetamidase marker was used once again to test its effectiveness in sequential deletions in *K. phaffii*. We observed ~600 transformants on MD medium containing acetamide, and 102 individual clones were transferred to MD medium plates containing uracil and 5-FOA to screen for the loss of *URA5* function. After 7 days of incubation at 28 °C, one colony that presented growth under these conditions was selected. Uracil auxotrophy was also confirmed in MD medium plates lacking uracil (data not shown), and the resulting strain was named LAU-*amdS*. Colony PCR using external confirmation primers URA5-F and URA5-R showed that the LAU-*amdS* strain had the *URA5* gene replaced by the *amdS* selection marker (Fig. 6).

One last marker recycling event was performed to remove *amdS* from the yeast genome. For reasons that are still unclear, this recycling step was not successful with the use of

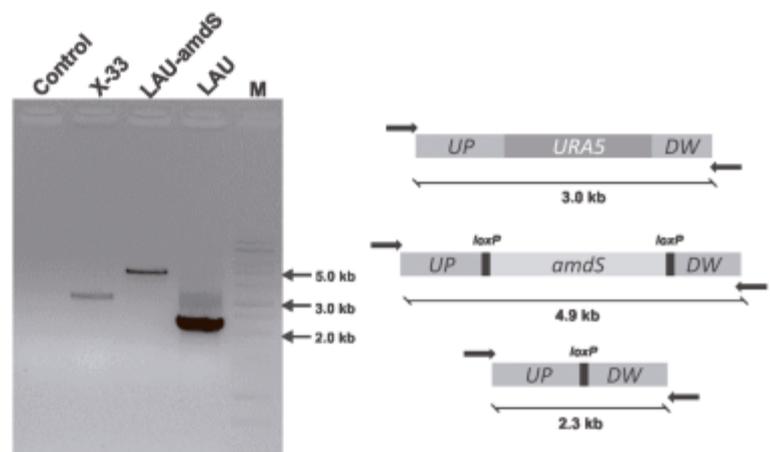
pYRCre2, which prompted us to construct a pPICH-Cre plasmid. Transformation of the LAU-*amdS* strain with pPICH-Cre required selection in YPDS supplied with zeocin, and clones were transferred to MD medium with fluoroacetamide for counter-selection. One clone that presented growth on MD medium with fluoroacetamide after 3 days of incubation was selected for colony PCR with primers URA5-F and URA5-R. Figure 6 shows that the resulting LAU strain provided a 2.3-kb amplicon while LAU-*amdS* yielded a 4.9-kb amplicon. The control X-33 strain with the wild-type *URA5* gene provided a 3.0-kb amplicon.

Finally, the *K. phaffii* strains generated by sequential gene deletion and marker recycling, LA2 and LAU, were plated on MD medium to confirm their auxotrophic phenotypes. As expected, both strains exhibited red colonies being LA2 Ade⁻ and LAU Ade⁻ and Ura⁻ (Fig. 7).

Discussion

The low availability of dominant markers represents an important bottleneck in *K. phaffii* strain development. This is particularly relevant when several genetic modifications are to be

Fig. 6 Deletion of *URA5* and marker recycling. Gel electrophoresis analysis of colony PCR products using primers URA5F and URA5R and different DNA templates. Lane 1: no template; lane 2: X-33; lane 3: LAU-*amdS*; lane 4: LAU; M: Promega 1 kb DNA Ladder. The panel on the right is a schematic representation of expected amplicon sizes



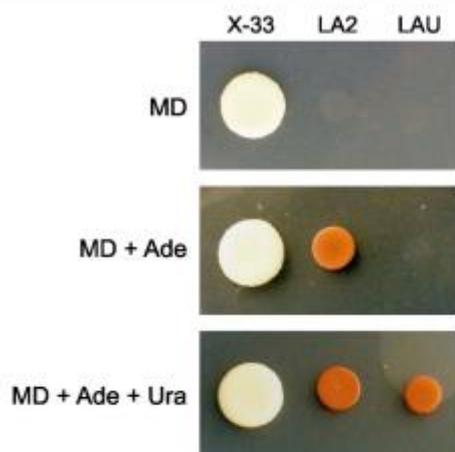


Fig. 7 Phenotypic analysis of strains LA2 and LAU. Cultures of *K. phaffii* X-33, LA2, and LAU were spotted on MD medium supplied or not with 0.002% adenine (Ade) or Ade + 0.2% uracil (Ura). Cells were grown at 28 °C for 4 days

performed in the same strain thus requiring marker recycling. The *A. nidulans amdS* gene is a dominant marker which offers a simple and relatively cheap selection of transformants through acetamide assimilation. Furthermore, it has the advantage of providing counter-selection on fluoroacetate.

When investigating acetamide assimilation by *K. phaffii*, we observed an endogenous acetamidase activity which we were able to link to *AMD2*, a gene related to other amidases and proteases described in literature (Kobayashi et al. 1997), which was shown to be constitutively expressed in all nitrogen sources tested. In *A. nidulans*, acetamidase activity is controlled by many regulatory genes influenced by both carbon and nitrogen sources: urea, ammonia, and glucose are involved in *amdS* repression, while acetamide and nitrogen-limiting conditions seem to enhance acetamidase production (Davis et al. 1993; Hynes 1970). A more thorough analysis will be required to uncover the role of *AMD2* in nitrogen metabolism in *K. phaffii*.

Both wild-type and an *amd2* mutant *K. phaffii* strain (LA1) were successfully transformed with a vector carrying *amdS* and a reporter gene. The fact that all colonies derived from LA1 represented true transformants demonstrates the effectiveness of *amdS* as a dominant marker in this strain. Since we could also recover a great number of transformants derived from X-33 cells, we conclude that it is possible to use *amdS* as a dominant marker in wild-type *K. phaffii* cells despite the small number of false positives that sporadically arise due to *AMD2* activity. Likewise, the fact that *S. cerevisiae* and *K. lactis* genomes also encode hypothetical amidase sequences sharing similarities to the *A. nidulans* acetamidase did not hamper the use of *amdS* as a selection marker in these yeasts (Chang and Abelson 1990; Ganatra et al. 2011; Solis-Escalante et al. 2013).

To further test *amdS*, we sequentially disrupted two *K. phaffii* genes, *ADE2* and *URA5*. The former was selected because adenine auxotrophs can be easily identified by colony color. Deletion of *ADE2* and/or *ADE1* cause cells to accumulate a red pigment (5-phosphoribosyl 5-aminoimidazole) while deletion of other genes in the adenine pathway, such as *ADE4* and *ADE5*, do not have the same effect and can even reverse the phenotype from red to white cells (Ugolini and Bruschi 1996). *URA5* mutants can be easily screened based on the counter-selection properties of this marker upon plating cells on 5-FOA (Boeke et al. 1984; Nett and Gemgross 2003).

Despite the high transformation efficiency, the low frequency of homologous integration events observed (3/~500 colonies for *ADE2* and 1/102 clones for *URA5*) can be explained by the effects of the non-homologous end joining mechanism (NHEJ) which is the preferred repair pathway in non-*Saccharomyces* yeasts (Kim et al. 2015). Many strategies have been developed in order to increase the efficiency of homologous recombination in *P. pastoris*, such as the creation of *ku70* mutant strains which are impaired in the NHEJ mechanism (Näätsaari et al. 2012). Split-marker approaches can increase homologous recombination rates by using the selection marker sequence to guarantee correct recombination, while a tailored CRISPR/Cas9 system provided targeting efficiencies close to 100% (Heiss et al. 2013; Weninger et al. 2016). Also, the success rate on integrating sequences using double integration techniques varies depending on the integration locus and the size of homologous sequences used for recombination (Li et al. 2007; Näätsaari et al. 2012). In a previous work, *URA5* was partially deleted in *K. phaffii* using a *kan* marker flanked by ~1 kb of homologous sequences on both sides, ~1.9% of the transformants were uracil auxotrophs (Nett and Gemgross 2003). Similarly, when we used the *amdS* selection marker and approximately the same size of homologous sequences for recombination, we obtained ~1% auxotrophic strains for uracil.

In between the gene deletion events, marker recycling also provided variable success rates. Transformation using plasmid pPICH-Cre provided better results in comparison to those obtained with pYRCre2 probably due to its native *K. phaffii* ARS1 sequence. Plasmid pYRCre2 contains a *S. cerevisiae* ARS1 sequence, which could lead to a less successful *K. phaffii* transformation since yeast replicating sequences have species-specific regions (Cregg et al. 1985).

In conclusion, *amdS* proved to be an effective selection marker for integrative transformation in *K. phaffii* also allowing easy marker recycling through counter-selection with fluoroacetamide. This counter-selection characteristic could prove useful in the construction of new strains that require several events of gene deletion or insertion, such as the new *K. phaffii* strains bearing heterologous pathways for metabolite production (Araya-Garay et al. 2012) or humanized glycosylation (Choi et al. 2003). Furthermore, recent studies

showed that the acetamidase selection marker used in *K. lactis* induced a higher frequency of multicopy integration when compared to antibiotic resistance markers (Read et al. 2007). In *K. phaffii*, such a characteristic would be of great use since a high gene copy number is often desired in order to improve protein productivity (Piva et al. 2017). Among the available yeast dominant selection markers, *amdS* thus represents an attractive addition to the *K. phaffii* genetic toolbox.

Acknowledgements This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq—Brazil) and Fundação de Apoio à Pesquisa do Distrito Federal (grant no. 193.000.582/2009).

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Apirion D (1965) The two-way selection of mutants and revertants in respect of acetate utilization and resistance to fluoroacetate in *Aspergillus nidulans*. *Genet Res (Camb)* 6:317–329

Araya-Garay JM, Feijoo-Siota L, Rosa-dos-Santos F, Veiga-Crespo O, Villa TG (2012) Construction of new *Pichia pastoris* X-33 strains for production of lycopene and β -carotene. *Appl Microbiol Biotechnol* 93:2483–2492

Betancur MO, Reis VCB, Nicola AM, De Marco JL, Moraes LMP, Torres FAG (2017) Multicopy plasmid integration in *Komagataella phaffii* mediated by a defective auxotrophic marker. *Microb Cell Factories* 16:99

Boeke JD, Lacroute F, Fink GR (1984) A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genomics* 197:345–346

Celik E, Calik P (2012) Production of recombinant proteins by yeast cells. *Biotechnol Adv* 30:1108–1118

Cereghino GPL, Cereghino JL, Ilgen C, Cregg JM (2002) Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Curr Opin Biotechnol* 13:329–332

Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24:45–66

Chang TH, Abelson J (1990) Identification of a putative amidase gene in yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res* 18:7180

Choi BK, Bobrowicz P, Davidson RC, Hamilton SR, Kung DH, Li H, Miele RG, Nett JH, Wildt S, Gerngross TU (2003) Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. *Proc Natl Acad Sci U S A* 100:5022–5027

Cregg JM, Barringer KJ, Hessler AY, Madden KR (1985) *Pichia pastoris* as a host system for transformations. *Mol Cell Biol* 5:3376–3385

Davis MA, Kelly JM, Hynes MJ (1993) Fungal catabolic gene regulation: molecular genetic analysis of the *amdS* gene of *Aspergillus nidulans*. *Genetica* 90:133–145

Delic M, Mattanovich D, Gasser B (2010) Monitoring intracellular redox conditions in the endoplasmic reticulum of living yeasts. *FEMS Microbiol Lett* 306:61–66

Drocourt D, Calmels T, Reynes J, Baron M, Tiraby G (1990) Cassettes of the *Streptoalloteichus hindustanus* ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Res* 18:4009

Fu J, Blaylock M, Wickes CF, Welte W, Mehrtash A, Wiederhold N, Wickes BL (2016) Development of a *Candida glabrata* dominant nutritional transformation marker utilizing the *Aspergillus nidulans* acetamidase gene (*amdS*). *FEMS Yeast Res* 16:fow023

Ganatra M, Vainauskas J, Hong JM, Taylor TE, Denson JPM, Esposito D, Read JD, Schmeisser H, Zoon KC, Hartley JL, Taron CH (2011) A set of aspartyl protease-deficient strains for improved expression of heterologous proteins in *Kluyveromyces lactis*. *FEMS Yeast Res* 11:168–178

Gomi K, Kitamoto K, Kumagai C (1992) Transformation of the industrial strain of *Aspergillus oryzae* with the homologous *amdS* gene as a dominant selectable marker. *J Ferment Bioeng* 74:389–391

Heiss S, Maurer M, Hahn R, Mattanovich D, Gasser B (2013) Identification and deletion of the major secreted protein of *Pichia pastoris*. *Appl Microbiol Biotechnol* 97(3):1241–1249

Hynes MJ (1970) Induction and repression of amidase enzymes in *Aspergillus nidulans*. *J Bacteriol* 103:482–487

Jørgensen MS, Skovlund DA, Johannesen PF, Mortensen UH (2014) A novel platform for heterologous gene expression in *Trichoderma reesei* (teleomorph *Hypocrea jecorina*). *Microb Cell Factories* 13:33

Kim H, Yoo SJ, Kang HA (2015) Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Res* 15:1–16

Kimura M, Kamakura T, Zhou Tao Q, Kaneko I, Isamu Y (1994) Cloning of the blasticidin S deaminase gene (BSD) from *Aspergillus terreus* and its use as a selectable marker for *Schizosaccharomyces pombe* and *Pyricularia oryzae*. *Mol Gen Genet* 242:121–129

Kobayashi M, Fujiwara Y, Goda M, Komeda H, Shimizu S (1997) Identification of active sites in amidase: evolutionary relationship between amide bond- and peptide bond-cleaving enzymes. *Proc Natl Acad Sci U S A* 94:11986–11991

Li P, Anumanthan A, Gao X-G, Ilangovan K, Suzara VV, Düzgüneş N, Renugopalakrishnan V (2007) Expression of recombinant proteins in *Pichia pastoris*. *Appl Biochem Biotechnol* 142:105–124

Lin Cereghino GP, Lin J, Jay A, Johnson MA, Lim M, Gleeson MAG, Cregg JM (2001) New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. *Gene* 263:159–169

Marx H, Mattanovich D, Sauer M (2008) Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*. *Microb Cell Factories* 7:23

Näätäsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A (2012) Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One* 7:e39720

Nett JH, Gerngross TU (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* 20:1279–1290

Nett JH, Hodel N, Rausch S, Wildt S (2005) Cloning and disruption of the *Pichia pastoris* ARG1, ARG2, ARG3, HIS1, HIS2, HIS5, HIS6 genes and their use as auxotrophic markers. *Yeast* 22:295–304

Piva LC, Betancur MO, Reis VCB, De Marco JL, Moraes LMP, Torres FAG (2017) Molecular strategies to increase the levels of heterologous transcripts in *Komagataella phaffii* for protein production. *Bioengineered* 8:441–445

Potvin G, Ahmad A, Zhang Z (2012) Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: a review. *Biochem Eng J* 64:91–105

Read JD, Colussi P, Ganatra MB, Taron CH (2007) Acetamide selection of *Kluyveromyces lactis* cells transformed with an integrative vector leads to high-frequency formation of multicopy strains. *Appl Environ Microbiol* 73:5088–5096

- Reis VCB, Nicola AM, Neto OSON, Batista VDF, de Moraes LMP, Torres FAG (2012) Genetic characterization and construction of an auxotrophic strain of *Saccharomyces cerevisiae* JP1, a Brazilian industrial yeast strain for bioethanol production. *J Ind Microbiol Biotechnol* 39:1673–1683
- Scorer CA, Clare JJ, McCombie WR, Romanos MA, Sreekrishna K (1994) Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. *New Biotechnol* 12:181–184
- Soderholm J, Bevis BJ, Glick BS (2001) Vector for pop-in/pop-out gene replacement in *Pichia pastoris*. *BioTechniques* 31:306–312
- Solis-Escalante D, Kuijpers NG, Bongaerts N, Bolat I, Bosman L, Pronk JT, Daran J-M, Daran-Lapujade P (2013) *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res* 13:126–139
- Sreekrishna K, Tschopp JF, Fuke M (1987) Invertase gene (*SUC2*) of *Saccharomyces cerevisiae* as a dominant marker for transformation of *Pichia pastoris*. *Gene* 59:115–125
- Sunga AJ, Cregg JM (2004) The *Pichia pastoris* formaldehyde dehydrogenase gene (FLD1) as a marker for selection of multicopy expression strains of *P. pastoris*. *Gene* 330:39–47
- Thor D, Xiong S, Orazem CC, Kwan A-C, Cregg JM, Lin-Cereghino J, Lin-Cereghino GP (2005) Cloning and characterization of the *Pichia pastoris* *MET2* gene as a selectable marker. *FEMS Yeast Res* 5:935–942
- Ugolini S, Bruschi CV (1996) The red/white colony color assay in the yeast *Saccharomyces cerevisiae*: epistatic growth advantage of white *ade8-18, ade2* cells over red *ade2* cells. *Curr Genet* 30:485–492
- Weninger A, Hatzl A-M, Schmid C, Vogl T, Glieder A (2016) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J Biotechnol* 235:139–149
- Yang J, Jiang W, Yang S (2009) *mazF* as a counter-selectable marker for unmarked genetic modification of *Pichia pastoris*. *FEMS Yeast Res* 9:600–609

Anexo 8. Artigo publicado com parte dos resultados obtidos no capítulo 1.

PLOS ONE

Functional characterization of *Komagataella phaffii* centromeres by a color-based plasmid stability assay –Manuscript Draft–

Manuscript Number:	PONE-D-18-31288R1
Article Type:	Research Article
Full Title:	Functional characterization of <i>Komagataella phaffii</i> centromeres by a color-based plasmid stability assay
Short Title:	Functional characterization of <i>Komagataella phaffii</i> centromeres
Corresponding Author:	Fernando Araripe Goncalves Torres, Ph.D. Universidade de Brasília Brasília, BRAZIL
Keywords:	<i>Komagataella phaffii</i> ; centromere; mitotic stability
Abstract:	The yeast <i>Komagataella phaffii</i> is widely used as a microbial host for heterologous protein production. However, molecular tools for this yeast are basically restricted to a few integrative and replicative plasmids. Four sequences that have recently been proposed as the <i>K. phaffii</i> centromeres could be used to develop a new class of mitotically stable vectors. In this work we designed a color-based genetic assay to investigate genetic stability in <i>K. phaffii</i> . Plasmids bearing each centromere and the ADE3 marker were evaluated in terms of mitotic stability in an <i>ade2/ade3</i> auxotrophic strain which allows plasmid screening through colony color. Plasmid copy number was verified through qPCR. Our results confirmed that the centromeric plasmids were maintained at low copy number as a result of typical chromosome-like segregation during cell division. These features, combined with high transformation efficiency and in vivo assembly possibilities, prompt these plasmids as a new addition to the <i>K. phaffii</i> genetic toolbox.
Order of Authors:	Luiza Cesca Piva Janice De Marco Lidia Moraes Viviane Reis Fernando Araripe Goncalves Torres, Ph.D.
Opposed Reviewers:	
Response to Reviewers:	<p>'Response to Reviewers'</p> <p>Reviewer #1: In the paper PONE-D-18-31288, entitled 'Genetic analysis of the <i>Komagataella phaffii</i> centromeres by a color-based plasmid stability assay', the authors have individually cloned 4 centromeres, previously identified in <i>K. phaffii</i>, on a shuttle replicative <i>E. coli</i> / yeast plasmid. They analyzed plasmid stability in the yeast over generations and quantified their copy number. They conclude that the centromeres provide mitotic stability and maintenance at low-copy number to the plasmids. The paper is clearly written. An added-value to the rather simple rationale is that, to test stability, the authors made smart use of a color assay based on partial complementation of <i>K. phaffii ade2 ade3</i> mutant by an ADE3 copy added on the plasmid. They also compare their results to recent publication by other groups, partially overlapping. I point out below some minor points.</p> <p>1- Improve descriptions:- Description of the exact centromere fragments used and what they contain could be improved by adding the genomic coordinates of the cloned fragments. This will help straightforward comparison with other publications. In particular, ARS were reportedly associated with some of the centromeres, are they present here? If so what could be the consequences of different ARS on the plasmid?</p> <p>ANSWER: We added the genomic coordinates regarding <i>K. phaffii</i> GS115 genome</p>

Functional characterization of *Komagataella phaffii* centromeres by a
color-based plasmid stability assay

Luiza Cesca Piva¹, Janice Lisboa De Marco¹, Lidia Maria Pepe de Moraes¹, Viviane
Castelo Branco Reis¹, Fernando Araripe Gonçalves Torres^{1*}

¹Laboratório de Biologia Molecular, Universidade de Brasília, Brasília, DF, Brazil,
70910-900.

*Corresponding author

E-mail: ftorres@unb.br

Abstract

The yeast *Komagataella phaffii* is widely used as a microbial host for heterologous protein production. However, molecular tools for this yeast are basically restricted to a few integrative and replicative plasmids. Four sequences that have recently been proposed as the *K. phaffii* centromeres could be used to develop a new class of mitotically stable vectors. In this work we designed a color-based genetic assay to investigate genetic stability in *K. phaffii*. Plasmids bearing each centromere and the *ADE3* marker were evaluated in terms of mitotic stability in an *ade2/ade3* auxotrophic strain which allows plasmid screening through colony color. Plasmid copy number was verified through qPCR. Our results confirmed that the centromeric plasmids were maintained at low copy number as a result of typical chromosome-like segregation during cell division. These features, combined with high transformation efficiency and *in vivo* assembly possibilities, prompt these plasmids as a new addition to the *K. phaffii* genetic toolbox.

Author summary

The methylotrophic yeast *Komagataella phaffii* is considered as one of the most important platforms for the production of proteins and metabolites. We sought in this study to develop a color-based genetic system widely used in other yeasts to assess mitotically stability of vectors carrying the proposed *K. phaffii* centromeres. First, we constructed a *K. phaffii* strain (LA3) mutant for *ADE2* and *ADE3*; this resulted in a strain that forms white colonies and when transformed with a vector (pPICH-*ADE3*) carrying *ADE3* turns red. Next, the four *K. phaffii* centromeres were cloned into pPICH-

ADE3 and tested in LA3 for copy number and plasmid stability. Centromeres are responsible for proper chromosome segregation during cell division, hence guaranteeing that both daughter cells receive one copy of the duplicated DNA. Our results show that three *K. phaffii* centromeres behaved as expected conferring extra stability to the replicative plasmids and maintaining them at low copy number. Once characterized, centromeres can be used as parts in the construction of advanced genetic manipulation tools, thus allowing the construction of strains capable of expressing large metabolic pathways for the production of complex biochemicals.

Introduction

Komagataella phaffii is a methylotrophic yeast of great industrial importance which has been used for more than 30 years as a heterologous protein production platform [1]. Its genome was first published in 2009 and has since then been refined and thoroughly studied [2,3]. As a result, in addition to a protein factory, *K. phaffii* has also been widely considered as a platform for the production of chemicals, biopharmaceuticals, vitamins and other molecules. However, the construction and regulation of new pathways demand complex molecular biology tools which are not readily available for this yeast [4].

K. phaffii genetic manipulation traditionally involves the use of shuttle vectors assembled in *Escherichia coli* and subsequently integrated into the yeast's genome [5]. Recent studies have described the development of a wide range of genetic parts for use in this yeast, as well as new methods of plasmid assembly and transformation [6]. An alternative to integrative strategies is the use of replicative plasmids, which are usually based on the well-known ARS1 sequence [1]. These plasmids may overcome some

drawbacks such as genetic instability in multi-copy strains and non-specific integration [7,8]. In addition, they present higher transformation efficiency when compared to integrative vectors and can be assembled by *in vivo* recombination, which eliminates the need for bacterial transformation [9,10]. However, replicative plasmids show low mitotic stability when compared to integrative vectors and few vector options are available for use [11]. Stability problems can be circumvented by the creation of centromeric plasmids, which may provide proper segregation during mitosis. A greater mitotic stability as well as low copy number allow stable and constant protein expression [12]. Centromeric plasmids can be constructed *in vivo*, allowing the assembly and cloning of large sequences including whole metabolic pathways and regulatory regions [13]. Therefore, the construction of such vectors would be of great value for *K. phaffii* strain development in the context of synthetic biology.

Centromeres are typically surrounded by large heterochromatin sections in most organisms [14]. Their structure ranges from simple “point” centromeres of only ~125 bp in *Saccharomyces cerevisiae* to epigenetic, sequence-independent centromeres, such as those present in plants and animals. The reason for this phenomenon is that, for most eukaryotes, centromeres are maintained epigenetically and not genetically. Sequence homologies are rare in and between species, hampering the definition of a consensus sequence. In addition, some DNA regions can be centromeric or not depending on its function in previous cell cycles, which highlights the epigenetic nature of the centromere [15].

As for non-conventional yeasts there are wide variations in centromere size and structure. *Candida glabrata* has centromeres that show some homology to the CDEI and CDEIII regions of *S. cerevisiae* while *Kuraishia capsulata* centromeres have 200-bp conserved sequences [16,17]. On the other hand, *Candida tropicalis*,

Schizosaccharomyces pombe and *Candida albicans* have regional centromeres named after their sizes which range from 3 to 110 kb [18–20].

K. phaffii centromeres have recently been identified, bearing no sequence similarities to those of any other yeast [3]. Since centromere function relies strongly on its structure rather than on its sequence, a centromere-specific histone H3 variant (CSE4) was used in the search for centromeric regions in *K. phaffii*. A CSE4 homolog was identified in chromosome 2 and tagged with a fluorescence marker. The corresponding nuclear localization of the histone-DNA complex indicated a centromere pattern typical of budding yeasts [3]. Tridimensional conformation analysis followed the centromere clustering pattern observed in yeasts and narrowed down all four *K. phaffii* centromere locations to 20 kb windows [21].

Considering that a low transcription rate is typical of centromeric regions, RNA-seq analysis allowed to pinpoint the putative centromeric locations for all four *K. phaffii* centromeres [3]. Similarly to *C. tropicalis* and *S. pombe*, *K. phaffii* centromeres are formed by inverted repeats. All four sequences have two inverted repeats of ~2,5 kb, separated by a central segment of 800 to 1300 bp. Chromatin immunoprecipitation sequencing analysis showed that the CSE4 histone binds preferably to the central region, but also along the inverted repeats [22].

K. phaffii centromeric sequences contain early replication peaks with autonomously replicating sequences, characteristics that are also observed in centromeres of other yeasts [23,24]. According to recently published studies, there are native ARS sequences contained within centromeres 2, 3 and 4. These comprise regions within the inverted repeats, as well as unique adjacent sequences [22,25].

In order to expand the functional analysis of the *K. phaffii* centromeres we sought in this study to develop a genetic system based on an *ade2/ade3* auxotrophic strain and a

replicative vector carrying the wild-type *ADE3*. Vectors carrying each individual centromere were used to assess plasmid copy number and mitotic stability.

Results and Discussion

In yeasts, adenine synthesis pathway is used as a tool for auxotrophic selection, gene copy number indicator and for plasmid stability analysis [26]. Many genes from this pathway have been deleted in *S. cerevisiae* in order to create auxotrophic strains, while in *K. phaffii* studies have only focused on *ADE1* and *ADE2* [27,28]. *K. phaffii* LA2, a strain mutant for *ADE2* [29], was used as a starting point for the construction of a strain that would allow plasmid stability verification. Deletion of *ADE2* results in cells auxotrophic for adenine which accumulate a red pigment [26] while deletion of genes located upstream, such as *ADE1* or *ADE3*, should prevent the formation of such pigment [27]. As expected, the deletion of *ADE3* in LA3 strain results in white colonies (Fig 1). Deletion of *ADE3* in *S. cerevisiae* has regulatory effects in the histidine synthesis pathway [30]. Consequently, *ade2 ade3* strains are not only auxotrophic for adenine, but also for histidine. In order to verify if this phenotype is applicable to *K. phaffii*, we plated strains X-33, LA2 and LA3 on MD medium without supplementation, comparing growth and colony color to cells plated on MD medium with adenine and histidine (Fig 1). LA3 strain displayed the expected histidine auxotrophy phenotype, showing that the adenine-histidine pathways in *K. phaffii* and *S. cerevisiae* have common characteristics.

Fig 1. Strain phenotypic analysis on defined media. *K. phaffii* X-33 (wild-type), LA2 and LA3 cultures were spotted on MD medium with or without supplementation.

In order to assess plasmid stability, we first constructed plasmid pPICH-ADE3 bearing the *ADE3* gene (Fig 2). When transformed with pPICH-ADE3, LA3 cells should return to being red and any changes on colony color would allow a simple screening of plasmid loss [26]. Although adenine auxotrophy has been explored for other purposes in *K. phaffii* [28], this particular color-based system has not yet been used for measuring plasmid stability in this yeast.

Fig 2. Map of vector pPICH-ADE3. ARS1 is an autonomously replicative sequence and *Sh ble* is the zeocin resistance marker. The *NotI* site was used for cloning *K. phaffii ADE3* gene.

pPICH-ADE3 was used for cloning all four *K. phaffii* centromeres. Since it revealed extremely difficult to amplify entire centromeric regions we designed a strategy to amplify centromeres in halves in order to reduce fragment size and to avoid primer annealing inside the inverted repeats (Fig 3). Amplified fragments exhibited in their ends overlapping regions that would allow recombination between each other and with vector pPICH-ADE3. Centromeric primer sequences were designed using *K. phaffii* GS115 genome sequence as reference [2]. The amplified regions corresponded to the following chromosomal coordinates: chromosome 1 position 1401429-1406917 (5488 bp); chromosome 2 position 1543739-1550657 (6918 bp); chromosome 3 position 2204800-2211493 (6693 bp) and chromosome 4 position 1703369-1709958 (6589 bp).

Fig 3. Strategy for amplification of *K. phaffii* centromeres. Schematic representation of a typical *K. phaffii* centromere. Inverted repeats are represented by green arrows. Primer annealing regions are shown by small arrows.

Centromeric sequences are known as early replication regions and according to recently published studies there are native ARS sequences contained within centromeres 2, 3 and 4 [11, 22]. Fig 4 shows the relative positions of the ARS sequences within and around the *K. phaffii* centromeres. In chromosome 2, ARS are located on coordinates 1543374-1543971 (597 bp) and 1549967-1551156 (1189 bp). These sequences were partially amplified in this work, containing 232 and 690 bp, respectively. As for chromosome 3, there is an ARS located on coordinates 2204369-2205185 (816 bp) which was also partially amplified (385 bp). Chromosome 4 has an ARS on coordinates 1703466-1704103 (637 bp) which was fully amplified, as well as a partially amplified ARS (840 bp) located on coordinates 1709118-1710114.

Fig 4. Relative positions of ARS sequences around centromeres 2, 3 and 4 of *K. phaffii*. Regions in gray represent centromeric sequences amplified in this work. Open rectangles indicate the relative positions of identified ARS sequences [11].

LA3 strain was individually transformed with pPICH-ADE3 and all four centromeric plasmids (pPICH-CEN1-4). Plasmids pPICH-CEN1, 2 and 4 were verified for autonomous replication by plasmid rescue in *E. coli*, while the circular structure of pPICH-CEN3 was confirmed through a set of overlapping PCRs since CEN3 was the only centromere that could not be cloned directly in bacteria.

Plasmid stability was firstly verified through colony color in non-selective medium (Fig 5). When plated on YPD non-selective medium, colonies transformed with pPICH-ADE3 lost their color rapidly and presented a red center with large white edges, a result consistent with plasmid instability. In contrast, strains transformed with all centromeric plasmids presented a uniform red coloration throughout the colony.

Fig 5. Plasmid stability analysis. LA3 strain transformed with pPICH-ADE3 and the four centromeric plasmids was grown on YPD medium for 3 days until color development.

Further stability examination of the centromeric plasmids was performed by growing cells in liquid YPD medium for 144 hours. After diluting and plating cultures on non-selective medium, red and white colonies were counted and compared between each construction (Fig 6). LA3 strain transformed with pPICH-ADE3 did not yield red colonies in any growth period, indicating that the plasmid was mitotically unstable. Conversely, centromeric plasmids presented a higher mitotic stability than pPICH-ADE3. After 96 hours of growth, cells with pPICH-CEN1 started to present white colonies, while all other centromeric plasmids remained stable. After 144 hours, pPICH-CEN1 was lost in most colonies while the other centromeric plasmids were lost in <10% cells. The reason for the instability of pPICH-CEN1 could be related to the absence of an autonomously replicating sequence within the centromere, since all other centromeres were cloned with at least a partially amplified ARS. The original replicating sequence in the pPICH-ADE3 plasmid, ARS1, has shown to be less efficient than its modern counterparts, therefore new ARS contained in the centromeres could have enhanced the mitotic stability of the centromeric plasmids [11].

Fig 6. Plasmid stability test. (A) Aliquots of the liquid cultures were collected at after 96 and 144 hours of growth and plated on YPD medium. Red colonies represent cells that maintained the *ADE3*-containing plasmid, while white colonies have lost it. Red portions of the bars represent red colonies; light pink bars represent white colonies. (B) A plate representing a typical result after 144 h growth.

Yeast centromeric plasmids knowingly have a higher mitotic stability under non-selective conditions than common replicative vectors since they are equally segregated between daughter cells and therefore provide a uniform culture of cells containing the plasmid [26]. A centromeric vector containing *K. phaffii* CEN2 has been constructed and it presented an enhanced stability when compared to a replicative plasmid [25]. In addition to *K. phaffii* and *S. cerevisiae*, centromeric plasmids have been developed for other yeasts such as *S. pombe*, *C. glabrata* and *Scheffersomyces stipitis* and in all cases enhanced plasmid stability under non-selective conditions was verified [12,17].

Yeast replicative plasmids are normally replicated but are unevenly distributed between daughter cells, which creates both multi-copy and plasmidless cells [26]. Under selective conditions, cells lacking the plasmid are unable to survive and the result is a population of multi-copy plasmid-containing cells. The construction of centromeric plasmids should provide better plasmid segregation and stability and cells should maintain a low and stable plasmid copy number during yeast growth [31]. Plasmid copy number was assessed by qPCR after strains were grown in YPD medium containing zeocin in order to ensure that all cells assayed were harboring the centromeric plasmids. The results were compared to LA3 strain transformed with pPICH-ADE3 also grown in selective medium and to the LA3 control strain, grown in YPD medium. Results from

qPCR (Fig 7) indicate that the strain transformed with centromeric plasmids carried 1-2 copies per cell while the replicative plasmid was present at approximately 25 copies per cell. The difference between plasmid copy number for the replicative vector and all centromeric vectors was significant according to a t-test ($p < 0.05$). This result illustrates the expected segregation pattern described above for growth in selective conditions and, together with the mitotic stability analysis, provides a clear picture of *K. phaffii* genetic manipulation using centromeric plasmids.

Fig 7. Plasmid copy number determination. The number of plasmids in each cell was estimated by qPCR. LA3 was used as negative control. Statistical analysis comparing each of the centromeric plasmids with the replicative vector was performed through a t-test using GraphPad Prism 5. ($p < 0.05$). Error bars depict the standard deviation of the mean ($n = 3$).

S. cerevisiae centromeric plasmids, in comparison to plasmids bearing the 2 μ m sequence, presented the same difference in copy number when auxotrophic markers were used. However, when the *kanMX* G418 resistance marker was used, plasmid copy number did not differ between centromeric and replicative plasmids [36]. This indicates that factors other than the type of replication origin can influence plasmid copy number. In a previous study, *K. phaffii* was transformed with replicative and integrative vectors bearing centromere 2 and, unlike our results, a low copy number of vector sequences was observed in both cases [25]. The reason for this discrepancy is unclear but it could be related to strain variation or plasmid constructions used since pPICH-CEN1-4 contained the *ADE3* gene in addition to ARS1 and the zeocin-resistance marker.

Overall, our results indicate that centromeric plasmids could be employed as a new tool for genetic manipulation of *K. phaffii*. Plasmids were maintained for long periods in non-selective medium, indicating that growth can be performed without the addition of antibiotics or any form of selective pressure. The centromeric plasmids' low copy numbers per cell characterize a stable and homogeneous culture that can provide reliable expression results. Finally, their structure as a circular molecule allows *in vivo* plasmid assembly with relatively short homologous sequences when compared to genomic integration techniques where sequences have to be much longer for directed homologous recombination. Simpler assembly may also facilitate the construction of larger and more sophisticated vectors such as yeast artificial chromosomes (YAC) whose stability features may be also analysed by the color-based assay described in this work.

Methods

Strains and Media

DNA cloning was performed using chemically competent *Escherichia coli* XL-10 Gold (Agilent Technologies) grown in LB medium (5 g L⁻¹ yeast extract, 10 g L⁻¹ peptone and 10 g L⁻¹ NaCl, pH 7,2). When needed, agar was added to a final concentration of 1,5%. When zeocin (25 µg mL⁻¹) was used for bacterial antibiotic selection, NaCl concentration was reduced to 5 g L⁻¹.

K. phaffii strains were derived from X-33 (Invitrogen). LA2 strain (*amd2 ade2*) was described in a previous work²⁹. Yeast was routinely grown in YPD medium (10 g L⁻¹ yeast extract, 20 g L⁻¹ peptone and 20 g L⁻¹ glucose). Solid medium used 2% agar. Zeocin and G418, when used, were added at 100 µg mL⁻¹ and 500 µg mL⁻¹,

respectively. Hygromycin B was used to a final concentration of 50 $\mu\text{g mL}^{-1}$. Minimal medium (MD) used 0,34% Yeast Nitrogen Base, 1% $(\text{NH}_4)_2\text{SO}_4$, 2% glucose, 0,00004% biotin and 0,0002% adenine or 0,004% histidine, when needed.

PCR

DNA was amplified using Invitrogen Platinum Taq DNA Polymerase (High Fidelity), Promega GoTaq Colorless Master Mix or Sigma-Aldrich Accutaq LA DNA Polymerase. All primers used in this work are shown in Table 1.

Table 1. Primers used in this work.

Primer	Sequence	Enzyme
ADE3up-F	GATAAGCTTGATATCGAATTCCTGCAGCCCCC GGGACGTAATGGAATAACTGCTGAC	<i>SmaI</i>
ADE3up-R	GAAGTTATGGATCCTACGAGGTAATTGAAGGCT CAC	
ADE3lox-F	CTTCAATTACCTCGTAGGATCCATAACTTCGTAT AATG	
ADE3lox-R	CAATCTCTCCCTTGTCATCGGATCCATAACTTCG TATAG	
ADE3dw-F	GAAGTTATGGATCCGATGACAAGGGAGAGATT GAAG	
ADE3dw-R	GGTGGCGGCCGCTCTAGAACTAGTGGATCCCC CCCGGGTGCAATGTACTGTTGAGTAGG	<i>SmaI</i>
ADE3conF	GGGACCGGAGGTAAAAGAC	
ADE3conR	GTTGGAATAATTGCATGGTCTG	
MUT1(Hpa)	CCATTGACATGGTAAACAGTTGGA	
MUT2(Bam)	GAACACTGGGATCTTGTTGAGG	
Cen1-F	GAGTTTAAACGGATCCACGAAGCAATGGATAG GCACT	<i>BamHI</i>
Cen1-R	CCGCGAATTCGGATCCTGAAGTCTTTCAGAGAG GAGCA	<i>BamHI</i>
Cen1c-F	CAAGTATGCGTGATCCCAGGT	

Cen1c-R	TACGAATTGTGGGGCTCTGT	
Cen2-F	GAGTTTAAAC <u>CGGATCC</u> ATCTCCGTTGATACTCC CAAC	<i>Bam</i> HI
Cen2-R	CCGCGAATTC <u>CGGATCC</u> ATCGACAAGCAGAACA CTAAG	<i>Bam</i> HI
Cen2c-F	GAATGGAGGTGCTGGTGGTTA	
Cen2c-R	TGTAATGCTCGCTGGTGAGT	
Cen3-F	GAGTTTAAAC <u>CGGATCC</u> AAGTGGTACACCAGTCA GCG	<i>Bam</i> HI
Cen3-R	CCGCGAATTC <u>CGGATCC</u> TCACTAGTATTCAACTGCAA CTGC	<i>Bam</i> HI
Cen3c-F	TCAGCCGAATACCCACACTT	
Cen3c-R	TCAGCCGTCAGCGAAATGAT	
Cen4-F	GAGTTTAAAC <u>CGGATCC</u> CAAACGCACCGTCTTGT TCA	<i>Bam</i> HI
Cen4-R	CCGCGAATTC <u>CGGATCC</u> AATTGATGTAGACGAGC AGC	<i>Bam</i> HI
Cen4c-F	TCAAGAATCGTACTGGCACCT	
Cen4c-R	CAAGCTCGTGAGATGGGATGT	
Cen370-F	CGCTCAGTGGAACGAAAACACGTTAAGGGA TTTGGTCATGAGATCAGATCTAACATCCAAAGA CGAAAGGTTGAATGAGTTTAAACGGATCCAAGT GG	
Cen370-R	GCTGGCCCTCTCTTCCCAGCTCACGAATCAGAT CCTAAGTCCTACTCAACAGTACATTGCAGCGGC CGCGTTTAAACGAATTCGGATCCTCAGTATTCA A	
qZEO-F	CGACGTGACCCTGTTCATCA	
qZEO-R	TGGACACGACCTCCGACCA	
qHIS-F	GTGTATCCTGGCTTGGCATCT	
qHIS-R	GCCAAGTACGGTGTGACGTT	

Restriction sites are underlined.

DNA manipulation

All basic DNA manipulation and analysis were performed as previously described [32]. Restriction digestion was performed in accordance to the manufacturer instructions (New England Biolabs), as well as vector dephosphorylation with Shrimp Alkaline Phosphatase (Promega) and ligation with T4 DNA ligase (USB). In-Fusion Cloning Kit (Clontech) was used for *in vitro* assembly of plasmids. Site-directed mutagenesis was performed using the Transformer Site-Directed Mutagenesis kit (Clontech). PCR and gel purification used Promega Wizard SV Gel and PCR Clean-Up System.

Quantitative PCR (qPCR)

Strains harboring the zeocin resistance plasmids were grown to an OD (optical density measured at 600nm) of 1 in 10 mL YPD containing zeocin while LA3 was grown in 10 mL YPD. Cells were collected by centrifugation at 2000 x g for 5 minutes. Cell pellet was resuspended with 1 mL 0,25% SDS and incubated at 98°C for 8 minutes according to a previous work [25]. Finally, cell debris was removed by centrifugation and DNA was diluted 10-fold in water before qPCR reactions.

Quantitative PCR reactions used primers qZEO-F and qZEO-R for plasmid quantification and qHIS-F and qHIS-R as an internal single-copy control. Assays were carried out with iTaq Universal SYBR Green Supermix (Bio-Rad) in a Rotor-Gene Q (Qiagen) thermal cycler. Analysis used the absolute quantification method and standard curves that ranged from 1×10^4 to 1×10^8 copies of the gene of interest. pPIC9 (Invitrogen) and pPICH linearized plasmids were used for construction of the standard curves.

Yeast transformation

K. phaffii was electroporated following two different protocols. For integrative cassettes, we followed the *Pichia* Expression Kit protocol (Invitrogen) and when using replicative plasmids, we proceeded as described previously [33].

Construction of an *ade2 ade3* strain for color-based stability assays

Strain LA2 [29] was transformed with an *ADE3* deletion cassette and had the marker recycled before moving on with the centromeric plasmid transformations. Construction of the deletion cassette used PCR reactions assembled by an “In-Fusion” cloning reaction. Briefly, primers ADE3up-F and R; ADE3dw-F and R were used for PCR amplification of 491 bp and 582 bp, respectively, from *K. phaffii* genome. These reactions amplified sequences used for directing homologous recombination and substitution of the complete *ADE3* coding sequence. Meanwhile, primers ADE3lox-F and ADE3lox-R amplified the *kan^R* G418 resistance cassette from plasmid pGKL [34]. PCR fragments were assembled and cloned into pBluescript II SK⁺ linearized with *Sma*I. A final PCR reaction using primers ADE3up-F and ADE3dw-R amplified the whole deletion cassette which was used for transformation of *K. phaffii* LA2. Cells were selected in YPD containing G418.

The resulting *ade2/ade3* strain was later transformed with pYRCre2 [35] and selected in YPD supplied with hygromycin B. This step promoted a Cre-mediated excision of the *kan* cassette thus eliminating G418 resistance. After PCR confirmation of marker recycling using primers ADE3conF and ADE3conR, the resulting strain was plated in non-selective YPD medium, causing loss of the pYRCre2 plasmid. The resulting strain was named LA3.

Construction of centromeric plasmids containing *ADE3*

Plasmid pPICH [29], which is derived from pPICHOLI (MoBiTec), contains the ARS1 replicating sequence [1]. This sequence is originally located on *K. phaffii* GS115 chromosome 2, coordinates 413701-413856 [2]. The plasmid was digested with *NotI* for cloning of the *K. phaffii* native *ADE3* gene. The complete gene was amplified from X-33 DNA using primers ADE3up-F and ADE3dw-R following digestion with *NotI*. After vector dephosphorylation, fragments were ligated and transformed into *E. coli* XL-10 Gold. One positive clone was then submitted to site-directed mutagenesis using primers Mut1(Hpa) and Mut2(Bam) for removal of the *Bam*HI restriction site present within the *ADE3* coding sequence. The final plasmid containing ARS1, the *Sh ble* resistance marker and *ADE3* was named pPICH-ADE3.

pPICH-ADE3 was digested with *Bam*HI for cloning of all four *K. phaffii* centromeres. These were amplified from *K. phaffii* X-33 genomic DNA using two PCR reactions for each centromeric sequence. Primers Cen1/2/3/4-F and Cen1/2/3/4c-R amplified the first inverted repeat of each centromere while primers Cen1/2/3/4c-F and Cen1/2/3/4-R amplified the other half of the sequences. In order to promote *in vitro/in vivo* assembly amplicons had approximately 80 bp homology between each other and 15 bp with pPICH-ADE3. Firstly, we attempted an “In-Fusion” cloning reaction for each of the four centromeres using linearized pPICH-ADE3 and the two PCR fragments. Plasmids were extracted and analyzed by restriction digestion. Centromeres 1, 2 and 4 were successfully assembled and cloned into the plasmid through this strategy.

Centromere 3 did not yield any *E. coli* clones following the “In-Fusion” reaction; therefore, we proceeded to an *in vivo* assembly strategy. Primers Cen370-F and Cen3c-R; Cen3c-F and Cen370-R amplified both inverted repeats adding 70 bp of

homologous sequences between the fragments and pPICH-ADE3. Finally, we transformed *K. phaffii* LA3 using the linearized vector and both centromeric fragments, using 85 bp of homology for directing recombination. Clones were selected in YPD supplied with zeocin.

The resulting plasmids were named pPICH-CEN1, pPICH-CEN2, pPICH-CEN3 and pPICH-CEN4. All plasmids were transformed into *K. phaffii* LA3 for subsequent stability and quantification assays.

Stability analysis

LA3 strain transformed with each of the four centromeric plasmids was grown in 20 mL YPD for 16 hours at 28°C and 200 rpm. This culture was inoculated to 20 mL YPD to an initial OD of 0,1. After 24 h of growth under the same conditions the culture was used as inoculum for another flask containing 20 mL YPD to an OD of 0,1. This procedure was repeated every 24 h until a total of 144 hours. At 96 and 144 hours of growth, a culture sample was diluted 10⁶-fold and 100 µL of this dilution were plated on YPD. Plates were incubated at 30 °C for 72 h.

References

1. Cregg, J. M., Barringer, K. J., Hessler, a Y., Madden, K. R. *Pichia pastoris* as a host system for transformations. *Mol. Cell. Biol.* 1985; 5(12): 3376–85.
2. De Schutter, K., Lin, Y.-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., et al. Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* 2009;27(6): 561–566.
3. Sturmberger, L., Chappell, T., Geier, M., Krainer, F., Day, K. J., Vide, U., et al. Refined *Pichia pastoris* reference genome sequence. *J. Biotechnol.* 2016; 235: 121–131.
4. Schwarzhans, J. P., Luttermann, T., Geier, M., Kalinowski, J., Friehs, K. Towards systems metabolic engineering in *Pichia pastoris*. *Biotechnol. Adv.* 2017;35(6): 681–

710.

5. Cereghino, J. L., Cregg, J. M. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* 2000; 24(1): 45–66.
6. Obst, U., Lu, T. K., Sieber, V. A modular toolkit for generating *Pichia pastoris* secretion libraries. *ACS Synth. Biol.* 2017; 6 (6): 1016-1025.
7. Camattari, A., Goh, A., Yip, L. Y., Tan, A. H. M., Ng, S. W., Tran, A., et al. 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(139). doi: 10.1186/s12934-016-0540-5
8. Schwarzhans, 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). doi: 10.1186/s12934-016-0486-7
9. Yu, X., Wang, R., Zhang, M., Xu, Y., Xiao, R. Enhanced thermostability of a *Rhizopus chinensis* lipase by in vivo recombination in *Pichia pastoris*. *Microb. Cell Fact.* 2012;11(102). doi : 10.1186/1475-2859-11-102
10. Lee, C. C., Williams, T. G., Wong, D. W. S., Robertson, G. H. An episomal expression vector for screening mutant gene libraries in *Pichia pastoris*. *Plasmid* 2005;54: 80–85.
11. Liachko, I., Youngblood, R. A., Tsui, K., Bubb, K. L., Queitsch, C., Raghuraman, M. K., et al. GC-Rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*. *PLOS Genet.* 2014;10:3. doi: 10.1371/journal.pgen.1004169
12. Cao, M., Gao, M., Lopez-Garcia, C. L., Wu, Y., Seetharam, A. S., Severin, A. J., et al. Centromeric DNA facilitates nonconventional yeast genetic engineering. *ACS Synth. Biol.* 2017; 6(8): 1545-1553.
13. Kouprina, N., Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* 2016; 125(4): 621-632.
14. Lyn Chan, F., Wong, L. H. Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Res.* 2012;40(22): 11178–11188.
15. Malik, H. S., Henikoff, S. Major evolutionary transitions in centromere complexity. *Cell* 2009;138(6): 1067–1082.
16. Morales, L., Noel, B., Porcel, B., Marcet-Houben, M., Hullo, M. F., Sacerdot, C., et

- al. Complete DNA sequence of *Kuraishia capsulata* illustrates novel genomic features among budding yeasts (Saccharomycotina). *Genome Biol. Evol.* 2013;5(12): 2524–2539.
17. Kitada, K., Yamaguchi, E., Arisawa, M. Isolation of a *Candida glabrata* centromere and its use in construction of plasmid vectors. *Gene* 1996;175(1-2): 105–108.
18. Ketel, C., Wang, H. S. W., McClellan, M., Bouchonville, K., Selmecki, A., Lahav, T., et al. Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. *PLOS Genet.* 2009;5(3). doi: 10.1371/journal.pgen.1000400
19. Chatterjee, G., Sankaranarayanan, S. R., Guin, K., Thattikota, Y., Padmanabhan, S., Siddharthan, R., et al. Repeat-associated fission yeast-like regional centromeres in the ascomycetous budding yeast *Candida tropicalis*. *PLOS Genet.* 2016;12(2). doi: 10.1371/journal.pgen.1005839
20. Smirnova, J. B., McFarlane, R. J. The unique centromeric chromatin structure of *Schizosaccharomyces pombe* is maintained during meiosis. *J. Biol. Chem.* 2002; 277(22); 19817–19822.
21. Varoquaux, N., Liachko, I., Ay, F., Burton, J. N., Shendure, J., Dunham, M. J., et al. Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic Acids Res.* 2015;43(11): 5331–5339.
22. Coughlan, A. Y., Hanson, S. J., Byrne, K. P., Wolfe, K. H. Centromeres of the yeast *Komagataella phaffii* (*Pichia pastoris*) have a simple inverted-repeat structure. *Genome Biol. Evol.* 2016;8(8): 2482–2492.
23. McCarroll, R. M., Fangman, W. L. Time of replication of yeast centromeres and telomeres. *Cell* 1988;54(4): 505–513.
24. Koren, A., Tsai, H., Tirosh, I., Burrack, L. S., Barkai, N., Berman, J. Epigenetically-inherited centromere and neocentromere DNA replicates earliest in s-phase. *PLOS Genet.* 2010;6(8). doi: 10.1371/journal.pgen.1001068
25. Nakamura, Y., Nishi, T., Noguchi, R., Ito, Y., Watanabe, T., Nishiyama, T., et al. Construction of a new stable, autonomously replicating plasmid vector containing *Pichia pastoris* centromeric DNA. *Appl. Environ. Microbiol.* 2018;84(15). doi: 10.1128/AEM.02882-17
26. Hieter, P., Mann, C., Snyder, M., Davis, R. W. Mitotic stability of yeast chromosomes : a colony color assay that measures nondisjunction and chromosome loss. *Cell* 1985;40(2): 381–392.
27. Ugolini, S., Bruschi, C. V. The red/white colony color assay in the yeast

Saccharomyces cerevisiae: epistatic growth advantage of white *ade8-18*, *ade2* cells over red *ade2* cells. *Curr. Genet.* 1996;30(6): 485–492.

28. Du, M., Battles, M. B., Nett, J. H. 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.

29. Piva, L. C., De Marco, J. L., de Moraes, L. M. P., Reis, V. C. B., Torres, F. A. G. Acetamidase as a dominant recyclable marker for *Komagataella phaffii* strain engineering. *Appl. Microbiol. Biotechnol.* 2018;102(6): 2753–2761.

30. Rébora, K., Laloo, B., Daignan-Fornier, B. Revisiting purine-histidine cross-pathway regulation in *Saccharomyces cerevisiae*: a central role for a small molecule. *Genetics* 2005;170(1): 61–70.

31. Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K., Verstrepen, K. J. Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiol. Rev.* 2014;38(5): 947–995.

32. Sambrook, J., Fritsch, E., Maniatis, T. *Molecular cloning: a laboratory manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press; 1989.

33. Wu, S., Letchworth, G. J. High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. *Biotechniques* 2004;36(1): 152–154.

34. Reis, V. C. B., Nicola, A. M., de Souza Oliveira Neto, O., Batista, V. D. F., de Moraes, L. M. P., Torres, F. A. G. 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(11): 1673–1683.

35. Betancur, M. O., Reis, V. C. B., Nicola, A. M., De Marco, J. L., de Moraes, L. M. P., Torres, F. A. G. Multicopy plasmid integration in *Komagataella phaffii* mediated by a defective auxotrophic marker. *Microb. Cell Fact.* 2017;16(99). doi: 10.1186/s12934-017-0715-8

36. Karim, A. S., Curran, K. A., Alper, H. S. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Res.* 2013;13(1): 107-116

Anexo 9. Artigo submetido com parte dos resultados obtidos no capítulo 2.

COMMENTARY



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, 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 ^{1,2}). 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.³ 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.⁴ The first expression system for *K. phaffii* was based on the promoter of the alcohol oxidase gene (P_{AOXI}) 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.^{5,6} P_{AOXI} 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.⁷ 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_{AOXI} .⁸ 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  Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF Brazil 70910-900.

[†]These authors contributed equally to this work.

© 2017 Taylor & Francis

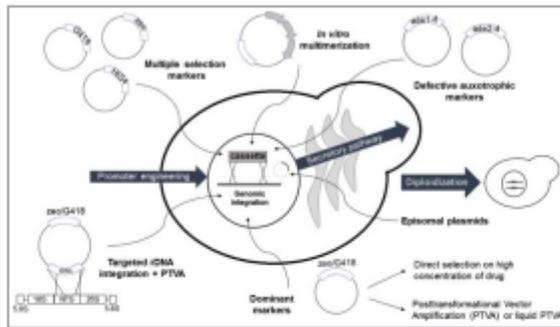


Figure 1. Molecular strategies used in *K. phaffii* to improve gene transcription. The titer of heterologous transcripts can be improved by using strong promoters (engineered or not) or by generating clones with extra doses of the desired gene integrated in the genome or present in episomal plasmids. 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.⁹ 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.¹⁰ 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.¹¹

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.¹² 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.^{13,14} In addition to P_{GAP} other moderately strong constitutive systems may be used such as P_{TEF1} ¹⁵ and P_{PGK1} .^{16,17}

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.⁴ 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.³ 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.¹⁸ 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.^{9,19} 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} .²⁰

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.^{3,21} In several cases this has led to a significant improvement in heterologous protein production, most notably when considering intracellular expression.²² 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.²³ 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.²⁴ Strategies to overcome these barriers may include the use of new signal peptides based on the *K. phaffii* secretome²⁵ and combinatorial engineering of secretion helper factors involved in protein folding and vesicle trafficking.²⁶

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

recombination and non-homologous end joining events.²⁷ Multiple integration events may occur at low frequency (5–6%) by the integration of expression cassettes *in tandem* in a head-to-tail configuration.²⁸ 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)²⁹ or its recently described variant “liquid PTVA”.³⁰ 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.²⁸ 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*.²²

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.^{3,22} Moreover, there are concerns that the accidental release of genetically modified organisms bearing drug-resistance markers may be horizontally transferred to environmental organisms.³¹ Marker-removal has been successfully accomplished in *K. phaffii* with the use of the Cre/loxP³² and Flp/FRT²⁷ site-specific recombination systems. The use of the *Escherichia coli* counter-selectable toxin gene *mazF* may further improve the screening of marker-free clones.³³ 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*³⁴ 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.^{35,36} 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.³⁷

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.³⁸ 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.³⁹

Recently, an episomal plasmid carrying the panARS sequence was tested for recombinant protein production in *K. phaffii*.⁴⁰ 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; <https://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; <https://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; <https://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; <https://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; [https://doi.org/10.1016/S0958-1669\(02\)00330-0](https://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; <https://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; <https://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; <https://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; <https://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; <https://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; [https://doi.org/10.1016/S0168-1656\(01\)00254-1](https://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; [https://doi.org/10.1016/S0378-1119\(96\)00675-0](https://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; <https://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; <https://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; <https://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; <https://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; <https://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; <https://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; <https://doi.org/10.1128/AEM.02843-10>
- [21] Romanos M, Scorer C, Sreekrishna 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; <https://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; <https://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; <https://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; <https://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; <https://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; <https://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; [https://doi.org/10.1016/0378-1119\(91\)90516-E](https://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; <https://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; <https://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; [https://doi.org/10.1016/S0168-1656\(98\)00105-9](https://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; <https://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; <https://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; <https://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; <https://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; <https://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; <https://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; <https://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; <https://doi.org/10.1186/s12934-016-0540-5>

Anexo 10. Artigo de revisão publicado ao longo deste e de outros trabalhos do laboratório de genética e biotecnologia de leveduras.