

GÉSSICA CABRAL BARROS

**EFEITOS DA DELEÇÃO DE *APJI* EM LEVEDURAS FERMENTATIVAS EM  
CONDIÇÕES DE ESTRESSE POR ETANOL**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Microbiologia Agrícola, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS - BRASIL  
2018

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

B277e  
2018

Barros, Géssica Cabral, 1992-  
Efeitos da deleção de *APJI* em leveduras fermentativas em  
condições de estresse por etanol / Géssica Cabral Barros. –  
Viçosa, MG, 2018.  
vi, 98f. : il. (algumas color.) ; 29 cm.

Inclui anexo.

Orientador: Luciano Gomes Fietto.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 70-77.

1. Leveduras. 2. Fermentação. 3. Biocombustíveis.  
4. Álcool. 5. Biotecnologia. I. Universidade Federal de Viçosa.  
Departamento de Microbiologia. Programa de Pós-Graduação  
em Microbiologia Agrícola. II. Título.

CDD 22. ed. 579.562

GÉSSICA CABRAL BARROS


**EFEITOS DA DELEÇÃO DE *AP11* EM LEVEDURAS FERMENTATIVAS EM  
CONDIÇÕES DE ESTRESSE POR ETANOL**

Dissertação apresentada à  
Universidade Federal de Viçosa,  
como parte das exigências do  
Programa de Pós-Graduação em  
Microbiologia Agrícola, para  
obtenção do título de *Magister  
Scientiae*.

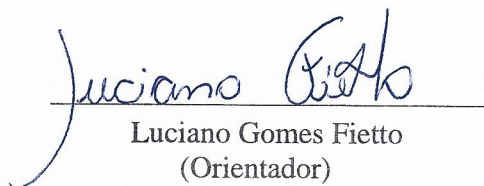
APROVADA: 29 de março de 2018.



Tiago Antônio de Oliveira Mendes  
(Coorientador)



Maximilân Dal-Bianco Lamas Costa



Luciano Gomes Fietto  
(Orientador)

## AGRADECIMENTOS

Primeiramente, gostaria de agradecer a Deus, por me dar forças e coragem para superar todos os obstáculos.

À Universidade Federal de Viçosa e ao Departamento de Microbiologia Agrícola pela oportunidade e conhecimento adquirido.

Aos meus pais, Gilmar e Marly, por todo amor, dedicação e por acreditarem em mim.

Aos meus irmãos, Gabriel e Luís Felipe, pelo carinho e companheirismo.

Ao restante da minha família pelo apoio e incentivo, especialmente tia Dê e tia Vanda.

Ao Professor Dr. Luciano Gomes Fietto pela orientação, conhecimento e apoio.

Ao Professor Dr. Tiago Antônio de Oliveira Mendes pela coorientação, disponibilidade e ensinamentos.

Aos amigos de laboratório, Alex, Patrícia, Gilza, Amanda e Fernanda, por todo conhecimento e ajuda nos experimentos, sem vocês com certeza eu não teria conseguido desenvolver este projeto.

A todos os amigos do LBM, pela convivência, risadas, prêmio Gilza, cafés, ajudas diversas e ensinamentos. Vocês fizeram meus dias mais felizes!

Aos meus amigos da Microbiologia, por toda ajuda nas disciplinas, conhecimento, companheirismo e momentos felizes.

À minha amiga de infância Thamyrys, pelo carinho, amizade e por me dar forças sempre que precisei.

Ao meu amigo Jessé, por todas as conversas e almoços juntos no ru.

Às amigas de república, por todas as conversas, apoio e incentivo.

Ao meu namorado Carlos, por todo amor, amizade, apoio, incentivo e paciência. Você foi essencial para esta conquista!

Aos professores do Departamento de Microbiologia e do Departamento de Bioquímica por todo o conhecimento transmitido.

Ao CNPq, FAPEMIG e CAPES, pelo apoio financeiro.

A todos que de alguma forma me ajudaram a chegar até aqui.

## SUMÁRIO

RESUMO .....	v
ABSTRACT .....	vi
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA .....	2
2.1 Produção de etanol e leveduras.....	2
2.2 Tolerância a etanol e <i>APJ1</i> .....	8
2.3 RNAseq: uma ferramenta para análises de transcriptomas.....	11
2.4 Sistema CRISPR-Cas9 e aplicações .....	13
3. OBJETIVOS .....	19
3.1 Objetivos Gerais .....	19
3.2 Objetivos Específicos .....	19
4. MATERIAL E MÉTODOS.....	20
4.1 Microrganismos e condições de cultivo.....	20
4.2 Tolerância a etanol das linhagens <i>S. cerevisiae</i> BY4741 e Y02999.....	20
4.3 Estratégia experimental- RNAseq .....	21
4.3.1 Extração de RNA total de <i>S. cerevisiae</i> Y02999 e BY4741 e análise da qualidade .....	21
4.3.2 Construção, validação e sequenciamento das bibliotecas .....	22
4.3.3 Tratamento dos dados e análises de bioinformática.....	24
4.3.4 Validação da expressão de alguns genes por qPCR.....	25
4.4 Estratégia experimental- sistema CRISPR-Cas9 .....	27
4.4.1 Desenho do gRNA e construção do vetor pML104-gRNA .....	27
4.4.2 Transformação de <i>K. marxianus</i> CCT7735 $\Delta$ <i>ura3</i> e <i>S. cerevisiae</i> AH109....	30
4.4.3 Validação da expressão de Cas9 em <i>K. marxianus</i> CCT7735 $\Delta$ <i>ura3</i> e <i>S.</i> <i>cerevisiae</i> AH109.....	31
4.4.4 Busca por transformantes positivos $\Delta$ <i>apj1</i> .....	32
5. RESULTADOS .....	34
5.1 Ensaio de tolerância ao etanol comparando <i>S. cerevisiae</i> Y02999 e BY4741 .....	34
5.2 RNAseq.....	36
5.2.1 Análise da qualidade dos RNAs extraídos .....	36
5.2.2 Análise das bibliotecas construídas.....	37
5.2.3 Análises de bioinformática.....	37
5.2.4 Análise dos genes escolhidos por qPCR .....	43
5.3 CRISPR-Cas9 .....	47

5.3.1 Clonagem do gRNA no pML104 .....	47
5.3.2 Validação da expressão de Cas9 .....	49
5.3.3 Busca por transformantes positivos $\Delta apj1$ .....	53
6. DISCUSSÃO .....	55
6.1 Influência da deleção de <i>APJ1</i> no transcriptoma de <i>S. cerevisiae</i> sob estresse por etanol.....	55
6.2 CRISPR-Cas9 utilizando o vetor pML104 em <i>K. marxianus</i> CCT7735.....	64
7. CONCLUSÃO.....	69
8. REFERÊNCIAS BIBLIOGRÁFICAS .....	70
<b>ANEXO A</b> .....	78

## RESUMO

BARROS, Gessica Cabral, M.Sc., Universidade Federal de Viosa, maro de 2018. **Efeitos da deleo de *APJ1* em leveduras fermentativas em condies de estresse por etanol.** Orientador: Luciano Gomes Fietto. Coorientador: Tiago Antonio de Oliveira Mendes.

O bioetanol e o principal biocombustivel utilizado mundialmente e e produzido atraves da fermentao de aucares por leveduras. A levedura *Saccharomyces cerevisiae* e a mais utilizada no processo, mas sua incapacidade na fermentao de pentoses levou ao estudo de outras especies que fermentam pentoses eficientemente, entretanto, estas especies apresentam baixa tolerancia a etanol. Assim, surgiu o interesse de avaliar, em nivel de transcriptoma, o papel do silenciamento de *APJ1* em *S. cerevisiae*, e promover sua deleo em *Kluyveromyces marxianus*, uma vez que trabalhos recentes indicam que sua deleo promoveu aumento da tolerancia a etanol. Duas linhagens de *S. cerevisiae*, BY4741 (*wild type*) e Y02999 ( $\Delta apj1$ ), foram utilizadas para estudo do transcriptoma na presenca de 8% (v/v) etanol. Apos um perodo de estresse de duas horas, foram identificados 61 genes *up*-regulados e 41 genes *down*-regulados em Y02999. Alem destes, 18 genes estavam presentes apenas em Y02999 e 138 genes somente em BY4741. Ao final, oito genes relacionados a tolerancia ao etanol e/ou envolvidos com o processo de sumoilao (funo relacionada a Apj1p) foram selecionados para validao da diferenca de expresso por qPCR, e os genes *ARG3* e *HSP104* se apresentaram *up*-regulados e o gene *OLE1* se apresentou *down*-regulado em Y02999 (12% v/v de etanol). Quanto a edio de *K. marxianus* CCT7735, o sistema CRISPR-Cas9 escolhido para adaptao foi o vetor pML04, desenvolvido para *S. cerevisiae*. A clonagem do RNA guia neste vetor no ocorreu da forma correta, entrando dois guias no stio de digesto. Alem disso, no foi possivel validar a expresso da Cas9 nesta levedura, uma vez que os diferentes *primers* desenhados para o experimento anelavam de forma inespecfica ao genoma da levedura no transformada. As colonias investigadas por sequenciamento que foram transformadas com esta construo, no apresentaram edio na sequncia de *APJ1*. Apesar do resultado negativo, a ferramenta CRISPR- Cas9 e promissora e cabem mais estudos e ensaios para padronizar a edio em *K. marxianus* CCT7735.

## ABSTRACT

BARROS, G ssica Cabral, M.Sc., Universidade Federal de Viosa, March, 2018. **Effects of APJI deletion in fermentative yeasts under ethanol stress conditions.** Adviser: Luciano Gomes Fietto. Co-adviser: Tiago Ant nio de Oliveira Mendes.

Bioethanol is the most consumed biofuel worldwide and is produced during the fermentation of sugars by yeasts. The yeast *Saccharomyces cerevisiae* is the most commonly employed organism in this process, however it is incapable of fermenting pentose, which promoted the prospection of other species that could efficiently ferment pentose, although these species present low ethanol tolerance. Therefore, there is increasing interest in the analysis of the role of APJI silencing in *S. cerevisiae* at transcriptome level and in the deletion of this chaperone in *Kluyveromyces marxianus*, given that recent studies indicate that this deletion resulted in the improvement of ethanol tolerance. Two strains of *S. cerevisiae*, BY4741 (wild type) and Y02999 ( $\Delta apj1$ ), were used for the transcriptome study, in the presence of ethanol 8% (v/v). After a two hours stress period, 61 up-regulated genes and 41 down-regulated genes were found in Y02999. Apart from these genes, 18 were present only in Y02999 and 138 were only in BY4741. Finally, eight genes related to ethanol tolerance and/or involved in the sumoylation process (function related to Apj1p) were selected to differential expression validation through qPCR. Genes *ARG3* and *HSP104* were up-regulated and *OLE1* was down-regulated in Y02999 (12% ethanol v/v). Regarding the *K. marxianus* CCT7735 edition, the vector pML04, which was developed for use in *S. cerevisiae*, was the CRISPR-Cas9 adaptation system of choice. The cloning of the RNA guide in this vector did not proceed as expected, since two guiding sequences were inserted in the digestion site. Furthermore, it was not possible to evaluate Cas9 expression in this yeast, once different primers designed for the experiment paired non-specifically with the untransformed yeast genome. The colonies transformed with this construction were analyzed by sequencing and did not present the edition in APJI sequence. Despite the negative result, CRISPR-Cas9 is a promising tool and more studies and assays should be done in order to establish a pattern for *K. marxianus* CCT7735 edition with this system.

## 1. INTRODUÇÃO

O etanol é o principal biocombustível utilizado no mundo, sendo produzido a partir de fontes renováveis de energia que substituem a utilização dos combustíveis fósseis poluentes como matéria-prima. O Brasil é o segundo maior produtor mundial de etanol, contribuindo significativamente para oferta mundial deste combustível.

A principal matéria-prima para produção de etanol no Brasil é a cana-de-açúcar, sendo que o processo de produção de etanol a partir da cana pode ser de primeira geração (1G) e segunda geração (2G). O etanol 1G é obtido diretamente da fermentação da sacarose presente no caldo da cana-de-açúcar, enquanto o etanol 2G é aquele obtido da fermentação de açúcares (hexoses e pentoses) presente em materiais lignocelulósicos, como o bagaço e a palha, permitindo um aumento na produção de etanol por unidade de hectare plantado.

O principal microrganismo utilizado para fermentação em escala industrial é a levedura *Saccharomyces cerevisiae*. O motivo desta preferência são as inúmeras vantagens que esta levedura apresenta em relação a outros microrganismos. Porém, esta levedura não tolera temperaturas muito elevadas e é incapaz de fermentar pentoses presentes no material lignocelulósico, o que limita a produtividade do etanol 2G.

Neste contexto, outras espécies de leveduras, que sejam capazes de contornar essas limitações supracitadas, despertam o interesse. É o caso das espécies *Kluyveromyces marxianus*, que tolera temperaturas elevadas, e *Scheffersomyces stipitis* e *Spathaspora passalidarum*, que são capazes de fermentar xilose. Infelizmente, a maior limitação do uso destas leveduras reside na sua eficiência de fermentação e tolerância ao etanol.

*APJ1* é um gene que se relaciona ao processo de tolerância ao etanol em linhagens de *S. cerevisiae*, sendo que sua deleção torna linhagens laboratoriais e industriais mais tolerantes a altas concentrações de etanol. Pouco se sabe sobre como a deleção deste gene leva ao aumento da tolerância e por isso, um dos objetivos deste trabalho foi avaliar o efeito da deleção de *APJ1* no transcriptoma de *S. cerevisiae* sob condição de estresse por etanol por meio da ferramenta RNAseq. Além disso, buscamos a deleção de *APJ1* em *K. marxianus*, através do sistema CRISPR-Cas9, com o objetivo de validar este sistema e avaliar o impacto da deleção de *APJ1* em outro organismo.

## 2. REVISÃO DE LITERATURA

### 2.1 Produção de etanol e leveduras

A substituição dos combustíveis fósseis por fontes renováveis de energia é um tema que tem se destacado nas últimas décadas em todo mundo. A queima de combustíveis fósseis para produção de energia libera gases (principalmente CO<sub>2</sub>) que contribuem para o efeito estufa, além daqueles que são nocivos à saúde humana (Junqueira *et al.*, 2017). As fontes renováveis de energia emergem como alternativas sustentáveis que diminuem os impactos da liberação de gases do efeito estufa e aquecimento global (Tesfaw *et al.*, 2014).

O etanol é o principal biocombustível utilizado no mundo (Kang *et al.*, 2014), sendo o Brasil o segundo maior produtor mundial ficando atrás somente dos Estados Unidos (EUA) (Figura 1). O etanol pode ser produzido a partir de várias plantas cultivadas, como milho, cana-de-açúcar, beterraba, mandioca e trigo. Os EUA utilizam o milho como matéria prima para produção e o Brasil utiliza a cana-de-açúcar, e juntos representam aproximadamente 80% da oferta mundial de etanol (Kang *et al.*, 2014).

Segundo informações da empresa Raízen<sup>1</sup>, o etanol produzido no Brasil a partir da cana apresenta um aproveitamento superior quando comparado ao etanol obtido do milho, pois o balanço energético do processo, ou seja, a diferença entre a energia gasta para converter a planta em etanol e quantidade de energia liberada ao final do processo, é cinco vezes maior. Outro fator que chama atenção para o etanol obtido a partir da cana, quando se considera a cadeia completa de produção, é que ele apresenta redução acentuada da emissão de gases do efeito estufa, em torno de 90% de emissão em comparação com a gasolina, enquanto para o milho e a beterraba, a redução é 30% e 45%, respectivamente (Amorim *et al.*, 2011).

O Brasil apresenta condições climáticas, solos férteis e tecnologias agrícolas que favorecem o desenvolvimento da cana-de-açúcar e conseqüentemente a produção de etanol (Amorim *et al.*, 2011), porém alguns fatores limitam o processo. Dentre eles, destacam-se: a colheita é limitada por 6-8 meses do ano e a cana não pode ser estocada, devido ao seu rápido processo de decomposição (Nova Cana<sup>1</sup>, Amorim *et al.*, 2011). Esses fatores podem ser o motivo do Brasil não ser o maior produtor mundial, tendo em vista que os EUA não sofrem estes problemas de colheita e estocagem, mantendo suas destilarias ativas quase o ano inteiro (Amorim *et al.*, 2011).

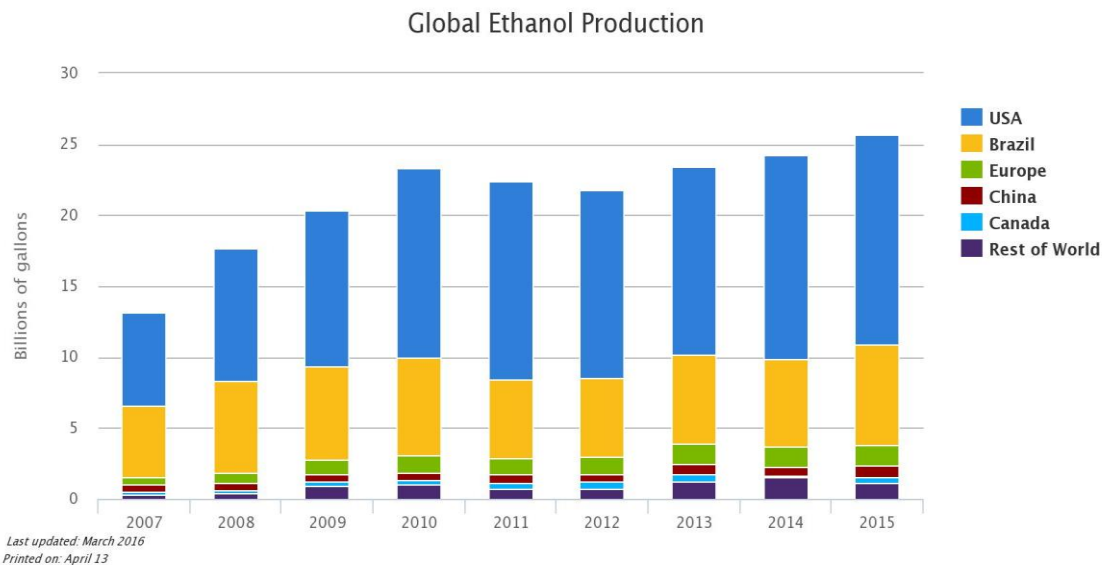


Figura 1: Produção de etanol em bilhões de galões no mundo Fonte: U. S. Department of Energy, Alternative Fuels Data Center, dados de 2015.

O etanol pode ser produzido através de dois processos: etanol de primeira geração (1G) e segunda geração (2G). O etanol 1G é aquele obtido diretamente da fermentação de açúcares e amido extraíveis da cana de açúcar e do milho, respectivamente (Junqueira *et al.*, 2017). O processo de produção de etanol 1G proveniente da cana-de açúcar inicia com a moagem da planta, onde o caldo é separado do bagaço. Em seguida, o caldo é tratado com temperatura elevada para eliminação de microrganismos contaminantes e hidrolisado para liberação de monômeros de glicose a partir da sacarose presente no caldo. O caldo é direcionado para dorna de fermentação, onde a glicose será convertida em etanol e CO<sub>2</sub> pela ação de leveduras, principalmente *Saccharomyces cerevisiae*. O líquido da fermentação é destilado para separar e purificar o etanol. O etanol então é desidratado para concentrações acima de 99,7%, para ser utilizado como combustível (Lennartsson *et al.*, 2014).

O etanol 2G é obtido a partir de diferentes tipos de materiais lignocelulósicos, como bagaço e palha da cana-de-açúcar, palha de soja, palha de arroz, dentre outros (Castro, 2010). O bagaço da cana é o principal resíduo agroindustrial brasileiro, apresentando grande potencial para geração de etanol 2G, sem precisar aumentar a área de plantio da cana-de-açúcar, permitindo a produção de etanol durante a entressafra da cana e aproveitando integralmente a planta (Nova Cana<sup>2</sup>).

Uma única unidade fabril da empresa Raízen destinada a produção de etanol 2G, tem expectativa de produção de 40 milhões de litros de etanol a mais por ano no Brasil

(Raízen<sup>2</sup>). Porém, o processo de produção ainda possui custos elevados e este custo precisa ser reduzido para mais da metade para ser economicamente sustentável (Souza *et al.*, 2011). O alto custo é explicado pelas diversas etapas adicionais, como o uso de enzimas para liberação de monômeros de açúcar livres para serem fermentados. Vários estudos estão sendo feitos para otimizar as etapas de produção e para o desenvolvimento de novas tecnologias, afim de tornar o processo viável economicamente (Souza *et al.*, 2011; Macrelli *et al.*, 2012).

A lignocelulose presente no bagaço da cana de açúcar é composta principalmente de celulose (40-60% do peso seco total), hemicelulose (20-40%) e lignina (10-25%) (Kang *et al.*, 2014). A celulose consiste em longas cadeias de glicose unidas por ligações glicosídicas do tipo  $\beta$ -1,4, organizadas em microfibras. A hemicelulose é formada por diferentes heteropolímeros incluindo hexoses (D-glicose, D-galactose, D-manose), pentoses (D-xilose e L-arabinose) e ácidos urônicos. A lignina é formada por compostos aromáticos, que conferem o alto nível de compactação e rigidez da parede celular de plantas. Juntos estes compostos formam uma matriz rígida de proteção da planta de difícil degradação, que precisa de vários tratamentos e enzimas para hidrólise, liberando as unidades monoméricas dos açúcares (Kang *et al.*, 2014).

As etapas de produção do etanol 2G consistem no pré-tratamento, sacarificação, fermentação e destilação (Aditiya *et al.*, 2016). Na etapa de pré-tratamento, ocorre a hidrólise da hemicelulose e solubilização da lignina para tornar a celulose mais acessível para as etapas posteriores (Aditiya *et al.*, 2016). Por isso, essa etapa é crítica para a eficiência de todo o processo. Um pré-tratamento ineficiente resultará em baixos rendimentos na produção de etanol 2G (Lu *et al.*, 2017). Tratamentos físicos, químicos, físico-químicos e biológicos são utilizados na despolimerização da estrutura da lignocelulose (Kang *et al.*, 2014).

Tratamentos físicos podem ser feitos através de redução mecânica, onde utilizam a moagem para diminuir o tamanho das partículas do material lignocelulósico e aumentar a área superficial para ação de enzimas nas etapas posteriores (Ogeda *et al.*, 2010). Tratamentos químicos, com ácidos concentrados ou diluídos (ácido sulfúrico ou ácidos carboxílicos), são utilizados para hidrólise da hemicelulose, enquanto a catalise básica (hidróxido de sódio ou amônia) ou uso de solventes orgânicos (etanol, metanol, propanol e acetona) ajudam na solubilização da lignina (Almeida *et al.*, 2011; Sarkar *et al.*, 2012). Tratamentos físico-químicos envolvem os processos de explosão a vapor, onde um vapor de água à alta pressão é gerado em um reator e age sobre o material lignocelulósico

durante um determinado tempo e em seguida sofre uma descompressão brusca, levando a separação dos constituintes do material (Pitarelo *et al.*, 2012). Tratamentos biológicos envolvem a utilização de enzimas produzidas por determinados microrganismos que são capazes de romper a estrutura da lignina e hemicelulose (Ogeda *et al.*, 2010).

A eficiência do pré-tratamento depende de diversos fatores determinados pela abordagem escolhida. Assim, o objetivo do pré-tratamento é obter altos rendimentos de separação dos constituintes do material lignocelulósico, evitar a formação de inibidores que comprometem as etapas de sacarificação e fermentação, e alcançar um processo economicamente viável. Os tratamentos mais utilizados são explosão a vapor e hidrólise com ácido diluído em temperaturas elevadas. A explosão a vapor é um tratamento com altos rendimentos de separação dos constituintes, baixa formação de inibidores, porém o custo do processo é elevado. Tratamentos com ácidos diluídos são preferencialmente utilizados no lugar de ácidos concentrados porque formam menos compostos inibidores (furfurais e hidroximetil-furfurais), apresentam alto rendimento de separação e são economicamente viáveis (Alvira *et al.*, 2010; Kang *et al.*, 2014).

A sacarificação corresponde a uma hidrólise enzimática para liberação de monômeros de glicose da celulose (Kang *et al.*, 2014). Os açúcares da hemicelulose também são interessantes para melhor aproveitamento do material lignocelulósico e maiores rendimentos na produção de etanol (Zeng *et al.*, 2017). Por isso, durante a sacarificação, um conjunto de enzimas hidrolisam a celulose e a hemicelulose e liberam as hexoses e pentoses para serem fermentados.

As celulasas formam um complexo enzimático que agem em sinergia para hidrolisar a celulose. Esse complexo é formado por três classes principais de enzimas: as endo-1,4- $\beta$ -D-glucanases são responsáveis por iniciar a hidrólise em regiões internas da celulose, clivando as ligações glicosídicas e gerando extremidades livres; as exo-1,4- $\beta$ -D-glucanases agem nas extremidades, hidrolisando os oligossacarídeos de glicose e gerando dissacarídeos (celbiose); as 1,4- $\beta$ -D-glucosidases hidrolisam a celbiose liberando glicose (Castro *et al.*, 2010; Ogeda *et al.*, 2010).

As hemicelulasas são enzimas responsáveis por romper a estrutura complexa e diversa da hemicelulose. Algumas das enzimas são: as endoxilanasas, que rompem internamente a estrutura da xilana; as exoxilanasas, que liberam xilbiose após a hidrólise de ligações glicosídicas  $\beta$ -1,4; as  $\beta$ -D-xilosidases, que hidrolisam a xilbiose e liberam a xilose. Além destas, muitas outras enzimas, como  $\alpha$ -L-arabinofuranosidase,  $\alpha$ -D-glucoranosidase, acetil-xilana-esterase, endo- $\beta$ -1,4-mannanase, exo- $\beta$ -1,4-mannosidase e

endo-galactanase são necessárias para liberar todos os açúcares presentes na hemicelulose (Shallom *et al.*, 2003).

O processo de sacarificação pode ocorrer separadamente ao processo de fermentação (SHF) ou pode ocorrer simultaneamente a fermentação (SSF). O processo SHF ocorre em reatores separados, ou seja: após o término da sacarificação, os açúcares livres são separados das enzimas utilizadas no processo e destinados a outro reator para serem fermentados por microrganismos específicos. A vantagem deste processo é a otimização das temperaturas da sacarificação e fermentação, tendo em vista que a atividade ótima das enzimas envolvidas na sacarificação é aproximadamente 50 °C e os microrganismos utilizados na fermentação crescem na faixa de 25-37 °C (Almeida *et al.*, 2011; Radecka *et al.*, 2015; Santos *et al.*, 2009). As maiores desvantagens são o alto custo para utilização de reatores separados e a inibição da atividade das enzimas da sacarificação pelo acúmulo de açúcares no meio (Kang *et al.*, 2014; Ogeda *et al.*, 2010).

O processo SSF apresenta as vantagens de diminuição de custos e ausência de inibições pelo produto da sacarificação, visto que o açúcar será fermentado em seguida a sua liberação no meio. A desvantagem do SSF é a dificuldade em se encontrar uma temperatura para o reator que possibilite a ocorrência eficiente tanto da sacarificação quanto da fermentação, sabendo que as faixas são distintas. Apesar da desvantagem, o rendimento e concentração de etanol são maiores no processo SSF (Ogeda *et al.*, 2010; Santos *et al.*, 2009).

A fermentação de hexoses é feita principalmente pela levedura *Saccharomyces cerevisiae*, por ser um microrganismo com capacidade de crescimento em condições aeróbicas e anaeróbicas, além de apresentar altas taxas de fermentação, alta tolerância a inibidores e etanol, quando comparada a outros microrganismos fermentadores, e tolerância a variações de pHs (Demeke *et al.*, 2013). Além disso, diversas linhagens de *S. cerevisiae* possuem o genoma sequenciado, o que permite a realização de manipulações genéticas de forma mais eficiente, e em escala comercial tem seu sistema de produção, estocagem e transporte bem estabelecidos (Demeke *et al.*, 2013; Radecka *et al.*, 2015; Tesfaw *et al.*, 2014). Porém, linhagens nativas de *S. cerevisiae* são incapazes de fermentar pentoses de forma eficiente, particularmente D-xilose, açúcar esse que corresponde a aproximadamente 35% do total de açúcares presentes na biomassa lignocelulósica (Demeke *et al.*, 2013). Por esta limitação, muitos pesquisadores se dedicam em encontrar linhagens de levedura que sejam capazes de fermentar pentoses para aumentar a produção de etanol 2G, e dentre as já isoladas, as espécies *Pachysolen tannophilus*, *Scheffersomyces*

*stipitis*, *Scheffersomyces shehatae* e *Spathaspora passalidarum* ganharam destaque pela habilidade em ter rendimentos relativamente altos de etanol a partir da xilose (Harner *et al.*, 2015). Porém, estas linhagens não-convencionais apresentam uma grande limitação para utilização em escala industrial, a baixa tolerância ao etanol quando comparadas a *S. cerevisiae* (Harner *et al.*, 2015).

*Kluyveromyces marxianus* é uma levedura bem conhecida por sua termotolerância, sendo capaz de crescer a 47 °C e produzir etanol em temperaturas acima de 40 °C. Também é capaz de utilizar xilose, arabinose, celobiose e xilitol como fontes de substrato. Essas características a tornam uma espécie interessante para o processo SSF de produção de etanol, pois consegue fermentar diversas pentoses presentes no material lignocelulósico e produzir etanol em temperaturas próximas a temperatura ótima das enzimas da sacarificação (Radecka *et al.*, 2015), o que diminui os custos de resfriamento do processo e evita contaminações com outros microrganismos (Diniz *et al.*, 2017). Outro ponto interessante desta levedura é a baixa repressão por glicose, ou seja, na presença de pentoses como fonte de carbono juntamente com glicose, ela consegue utilizar ambas as fontes de carbono simultaneamente para produção de etanol (Lertwattanasakul *et al.*, 2013).

Na tentativa de resolver as dificuldades do processo de produção de etanol, a grande maioria das pesquisas existentes na área tem focado em manipulações genéticas em uma única linhagem de levedura, para que ela possa ser eficiente na fermentação de hexoses e pentoses da biomassa lignocelulósica, tolerar a presença de inibidores, tolerar altas concentrações de etanol, ser termotolerante (para utilização no processo SSF), dentre outras características (Liang *et al.*, 2013).

Exemplo dessa modificação genética é a realização de engenharia metabólica em *S. cerevisiae* para expressão de enzimas heterólogas do metabolismo de xilose, como xilose redutase e xilitol desidrogenase ou xilose isomerase, visando torna-la capaz de fermentar essa pentose (Almeida *et al.*, 2011). Senatham *et al.* (2016) obtiveram um mutante de *Scheffersomyces shehatae* TTC79, através de mutagênese por exposição a luz UV, que foi capaz de fermentar glicose tão eficientemente quanto a linhagem *wild-type* e capaz de aumentar 58,95% seu rendimento na produção de etanol durante a fermentação de xilose. Além disso, este mutante apresentou uma melhora significativa na tolerância a inibidores presentes em hidrolisados lignocelulósicos, se tornando um potencial microrganismo para produção de etanol a partir de hidrolisados.

Hou *et al.* (2012) mostraram o aumento na tolerância a presença de inibidores pela levedura *Spathaspora passalidarum* - levedura capaz de fermentar glicose e xilose simultaneamente em condições de anaerobiose - utilizando uma combinação de mutagênese e fusão de protoplasto para construção das linhagens mutantes. O mutante *S. passalidarum* M7 obtido por mutagênese produziu 50% a mais de etanol comparado a linhagem *wild-type* em um meio contendo xilose e 2 g/L de furfural. Através da fusão de protoplasto do mutante *S. passalidarum* M7 e uma linhagem robusta de *S. cerevisiae* ATCC 96581, foi obtido um híbrido capaz de crescer em um meio contendo 75% da fração líquida do pré-tratamento da palha de trigo, onde diversos inibidores estão presentes.

Através da metodologia de “*Genome Shuffling*” entre as leveduras *S. stipitis* CBS 6054 e *S. cerevisiae* ATCC 24860, Zhang *et al.* (2012) construíram linhagens que apresentaram melhora na produção de etanol, tolerância a etanol e taxas mais rápidas de utilização de xilose. Uma linhagem obtida foi a ScF2, que durante a fermentação na presença de 50 g/L de glicose e 50 g/L de xilose foi capaz de utilizar os dois açúcares e produzir 40 g/L de etanol, enquanto a *S. cerevisiae* ATCC 24860 foi capaz somente de fermentar a glicose, com produção de etanol de 22 g/L, e a *S. stipitis* CBS 6054 que, apesar de fermentar os dois açúcares, apresentou uma menor produção de etanol (31 g/L) em relação a ScF2.

Diversos outros trabalhos envolvendo melhoramento genético de leveduras para produção de etanol são encontrados, indicando a importância do desenvolvimento de linhagens capazes de aproveitar ao máximo o material a ser fermentado e aumentar a produção de etanol 1G e 2G.

## 2.2 Tolerância a etanol e APJI

O rendimento da produção de etanol durante a fermentação nas condições industriais depende, dentre outros fatores, da viabilidade e capacidade do microrganismo utilizado em converter açúcar em etanol (Stanley *et al.*, 2010). Nos últimos estágios do processo, o acúmulo de etanol se torna tóxico para as leveduras, inibindo o crescimento celular e a viabilidade, consequentemente limitando a produtividade da fermentação (Snoek *et al.*, 2016). Diferentes linhagens de levedura apresentam diferenças em sua habilidade de crescer na presença do etanol, sendo que aquelas que apresentam maior tolerância são mais interessantes para uso industrial (Voordeckers *et al.*, 2015). Por estes motivos, é de grande interesse dos pesquisadores entender os mecanismos que levam a tolerância nos

microrganismos mais adaptados (por exemplo, *Saccharomyces cerevisiae*) e desenvolver estratégias para aumentar a tolerância e a capacidade de proliferação na presença do etanol (Pais *et al.*, 2013). Além disso, aplicar estas estratégias em leveduras que são pouco tolerantes, mas que apresentam outras características interessantes para o processo, é uma abordagem que permitiria o desenvolvimento de novas linhagens com competências específicas (Yuan *et al.*, 2011).

A toxicidade do etanol para as células é caracterizada pela influência deste no metabolismo celular e biossíntese de macromoléculas, diminuição dos níveis de proteínas, por afetar o dobramento e estabilidade destas, diminuição dos níveis de RNA mensageiro e redução da atividade de enzimas da via glicolítica (Stanley *et al.*, 2010). O etanol também altera a estrutura da mitocôndria, diminuindo a respiração, síntese de ATP e levando a formação de espécies reativas de oxigênio (ROS). A presença de ROS causam danos na estrutura do DNA cromossomal e mitocondrial (Navarro-Tapia *et al.*, 2017).

A membrana plasmática é uma das principais estruturas afetadas pelo etanol, comprometendo algumas funções celulares (Aguilera *et al.*, 2006; Stanley *et al.*, 2010). Um dos efeitos do etanol na membrana é o aumento da fluidez, que resulta no aumento da permeabilidade da membrana a prótons. O influxo de prótons resulta em rápida dissipação do gradiente eletroquímico, acidificação intracelular, perda da força próton-motiva e da atividade H<sup>+</sup>-ATPase. A perda de atividade H<sup>+</sup>-ATPase compromete diversos transportes através da membrana que são dependentes do gasto de energia, como a captação de glicose e aminoácidos (Aguilera *et al.*, 2006; Alexandre *et al.*, 2001; Navarro-Tapia *et al.*, 2017).

Na tentativa de sobrevivência ao estresse, as leveduras desenvolvem vários tipos de respostas e mecanismos de adaptação (Alexandre *et al.*, 2001). O processo de tolerância é complexo, envolvendo a reprogramação das atividades celulares no contexto de modulação da expressão gênica, sinalização celular, transdução de sinais, controle transcricional e pós-transcricional, acúmulo de moléculas protetoras e aumento das atividades de reparo das células (Stanley *et al.*, 2010). Alguns mecanismos estudados e relacionados a resistência à presença de etanol são: a alteração da composição dos fosfolipídeos da membrana para diminuir a permeabilidade ao etanol; acúmulo de trealose e ergosterol, que aumenta estabilidade de proteínas e da membrana (Teixeira *et al.*, 2009); indução de chaperonas e proteínas de choque térmico (HSPs), envolvidas no remodelamento de proteínas desnaturadas (Stanley *et al.*, 2010).; ativação da via de resposta a proteínas não dobradas (UPR, *unfolded protein response*) ( Navarro-Tapia *et*

*al.*, 2017); *up*-regulação de genes relacionados a regeneração de NADH/NADPH para manter o balanço redox nas células (Navarro-Tapia *et al.*, 2017); aumento da atividade da superóxido dismutase mitocondrial para eliminação de ROS (Teixeira *et al.*, 2009) e necessidade da presença de genes relacionados a montagem da H<sup>+</sup>-ATPase do vacúolo e da membrana plasmática, para manutenção da homeostase do pH intracelular (Charoenbhakdi *et al.*, 2016).

Diversas estratégias são utilizadas pelos pesquisadores para encontrar linhagens mais resistentes. Uma das estratégias é a utilização da técnica “*Genome Shuffling*” aplicada em larga escala, que permite combinar alelos de diferentes linhagens que são naturalmente tolerantes, para obter híbridos ainda mais tolerantes (Snoek *et al.*, 2016). Pode-se também investigar o efeito da deleção ou superexpressão de um único gene na tolerância ao etanol ou manipular simultaneamente múltiplos genes, introduzindo randomicamente cópias com mutações de um fator de transcrição que controla um amplo número de genes, por exemplo (Zhao *et al.*, 2009).

Outras abordagens úteis são: análise de transcriptoma e proteoma em células submetidas ao estresse etanólico e comparar o perfil de expressão gênica entre linhagens tolerantes e não tolerantes, objetivando encontrar alvos para manipulações genéticas. Estratégias de adaptações fisiológicas, como a evolução dirigida em linhagens não tolerantes, submetendo-as em meios com concentrações crescentes de etanol para obtenção de células mais tolerantes; e a utilização de mutações randômicas, por exposição a agentes mutagênicos, para obtenção de linhagens mais resistentes, dentre outras estratégias (Caspeta *et al.*, 2015).

Swinnen *et al.* (2012), através de análises do genoma de linhagens laboratoriais e industriais de *S. cerevisiae* envolvidas com a produção de etanol, descobriram possíveis genes que estão relacionados a tolerância em altas concentrações de etanol. As linhagens utilizadas no estudo são as linhagens VR1-5B, que é uma segregante da linhagem VR1 utilizada para fermentação em indústrias brasileiras, e BY4741, que corresponde a linhagem laboratorial. A segregante VR1-5B apresenta crescimento em meio contendo até 18% de etanol, enquanto a BY4741 apresenta pouco crescimento em meio sólido com 14% de etanol. Foi feito um cruzamento entre a linhagem segregante e a linhagem laboratorial e a partir dos segregantes obtidos foi feito um *pool* de segregantes que toleravam pelo menos 16% de etanol e outro *pool* que tolerava 17% de etanol. Os *pools* de segregantes e as linhagens parentais (VR1-5B e BY4741) foram submetidas ao sequenciamento do genoma e, a partir da análise da presença de polimorfismos de

nucleotídeo único (SNPs), encontrou-se alguns genes relacionados a tolerância a etanol e dentre eles o gene *APJI*. Durante o trabalho foi visto que *APJI* correlaciona-se de forma negativa ao processo de tolerância, sendo mais expressa em BY4741, que é a linhagem menos tolerante. A deleção de *APJI* na linhagem BY4741 fez com que o crescimento aumentasse em meio sólido contendo 14% de etanol, e a linhagem VR1-5B com a deleção apresentou uma pequena melhora no crescimento em todas as concentrações de etanol testadas.

Apj1p é uma proteína conservada entre os eucariotos e é caracterizada como uma chaperona putativa da família Hsp40p/DNAJ, apresentando um domínio J altamente conservado. Em *S. cerevisiae*, desempenha um papel na degradação de proteínas sumoiladas (*Saccharomyces Genome Database*) e tem localização mitocondrial (Swinnen *et al.*, 2012). Alexandre *et al.* (2001) realizaram um estudo sobre a resposta transcricional de *S. cerevisiae* S288c sob uma condição de estresse etanólico, onde as células foram submetidas ao tratamento com 7% de etanol durante 30 minutos, e dentre os genes induzidos durante o estresse ele encontrou *APJI* (*YNL077W*). Schwartz *et al.* (2012) viram a participação de um homólogo de *APJI* no metabolismo de xilose em leveduras do gênero *Saccharomyces* que são naturalmente capazes de metabolizar essa pentose. Poucas são as informações na literatura sobre *APJI* e suas funções biológicas em leveduras de interesse para produção de etanol. Mais estudos são necessários para entender sua correlação negativa ao processo de tolerância ao etanol em *S. cerevisiae* e aplicabilidade dessa deleção em outros microrganismos pouco tolerantes.

### 2.3 RNAseq: uma ferramenta para análises de transcriptomas

Transcriptoma é definido como o conjunto total de transcritos presentes em uma célula ou em uma população de células, em um estágio específico do desenvolvimento ou condição fisiológica (Wang *et al.*, 2009). Possui um alto nível de complexidade e é constituído por várias espécies de RNAs codificadores (RNAs mensageiros), não-codificadores e pequenos RNAs, que desempenham funções em diversos processos biológicos (Kukurba *et al.*, 2015). O estudo do transcriptoma é essencial para a identificação dos elementos funcionais do genoma, entendimento da variação da expressão gênica em cada condição que a célula for submetida e observação da ocorrência de *splicing* alternativos e modificações pós-transcricionais, quando existentes (Wang *et al.*, 2009).

Atualmente, a ferramenta mais utilizada pelos pesquisadores para análises de transcriptomas é o *RNA sequencing* (RNAseq) (Chu *et al.*, 2012). O RNAseq é uma técnica altamente sensível e precisa, que se baseia na tecnologia de sequenciamento de alta performance, *Next Generation Sequencing* (NGS), para revelar a identidade da maioria das espécies de RNA presentes na célula e permitir uma análise e quantificação mais detalhada da expressão gênica (Chu *et al.*, 2012; Kukurba *et al.*, 2015). Além disso, permite a detecção de outras modificações como a descoberta de novos transcritos, fusão gênica, variações em um único nucleotídeo e *splicing* alternativos (Wang *et al.*, 2009).

Um experimento típico de RNAseq começa com o isolamento do RNA total a partir das amostras biológicas que se deseja estudar, sendo que este RNA deve apresentar uma alta qualidade (Kukurba *et al.*, 2015). Em seguida, é feita a construção de uma biblioteca de DNAs complementares (cDNAs). Primeiramente, seleciona-se os RNAs de interesse, a partir da remoção dos RNAs ribossomais ou pela captura de RNA mensageiros a partir da cauda poli(A). Estes RNAs de interesse são fragmentados e, pela ação da transcriptase reversa, são convertidos em cDNAs. Às extremidades dos cDNAs de cada amostra biológica são ligados diferentes adaptadores (sequências curtas de nucleotídeos conhecidas), que permitem o sequenciamento de múltiplas amostras em uma mesma reação (Chu *et al.*, 2012, Wang *et al.*, 2009).

A principal plataforma de sequenciamento utilizada para realização do NGS para RNAseq é a Illumina HiSeq (McGettigan, 2012). Durante o sequenciamento nesta plataforma, acontecem vários ciclos de amplificação dos fragmentos, gerando um grande número de *reads*. Essas *reads* são curtas, com tamanhos em torno de 250 pares de bases. Esse tamanho varia dependendo do tipo de ligação dos adaptadores (uma ou duas extremidades dos fragmentos de cDNA) e da versão usada do HiSeq (Goodwin *et al.*, 2016).

Para transformar a grande quantidade de dados gerada no sequenciamento em um transcriptoma montado e obter os dados de expressão gênica, é necessária a utilização de diferentes ferramentas de bioinformática e estatística, sendo cada passo analisado por um conjunto específico de programas. O primeiro passo é avaliar a qualidade das *reads*, a fim de identificar anormalidades durante a preparação da biblioteca ou sequenciamento (Verk *et al.*, 2013). Em seguida, é feita a montagem do transcriptoma, que pode ocorrer de duas maneiras: quando se estuda um organismo que já possui o genoma sequenciado, a montagem é feita pelo alinhamento das *reads* a um genoma de referência ou transcriptoma de referência; em casos de estudos com organismos que não possuem genomas

sequenciados, é feita a montagem *de novo* das *reads* em *contigs* para obtenção das seqüências (Conesa *et al.*, 2016, Wang *et al.*, 2009).

Após a montagem, é feita a contagem de quantas *reads* foram mapeadas em cada gene, transformando os dados de alinhamento em dados quantitativos, e realiza-se uma normalização para tornar os dados de contagem comparáveis para detecção de diferenças de expressão. O último passo é a identificação dos genes que estão diferentemente expressos de forma significativa nas condições testadas no experimento através da utilização de programas estatísticos (McGettigan, 2012, Verk *et al.*, 2013). Em todos os passos da análise, é essencial fazer um controle de qualidade dos dados para garantir a reprodutibilidade e confiabilidade dos resultados obtidos (Conesa *et al.*, 2016).

O RNAseq é aplicado para diversos estudos e os resultados obtidos oferecem um panorama global da alteração do padrão de transcrição em um organismo sob uma dada condição. Feng *et al.* (2013) utilizaram essa ferramenta para estudar o metabolismo de xilose em duas linhagens de *Saccharomyces cerevisiae* que foram geneticamente desenvolvidas para utilizar essa pentose e produzir etanol. Em um estudo anterior, três diferentes vias da utilização de xilose foram introduzidas nestas duas linhagens e observou-se que linhagens diferentes apresentavam rendimentos de produção de etanol diferentes. Isso sugeriu aos pesquisadores que a produção de etanol a partir da xilose não estava apenas relacionada a via pela qual a levedura utilizava o açúcar, mas dependia também da linhagem que sofria as manipulações genéticas. Através do RNAseq, os pesquisadores puderam comparar a resposta transcricional das linhagens de *S. cerevisiae* com as diferentes vias e entender quais eram os genes envolvidos nas diversas estratégias de regulação do metabolismo em cada linhagem para aumentar a utilização da xilose e identificar os genes chaves que levavam a dependência da linhagem que recebia as vias.

#### 2.4 Sistema CRISPR-Cas9 e aplicações

O sistema CRISPR-Cas é um mecanismo natural de defesa de bactérias e arqueas contra um DNA invasor (podendo ser material genético viral, plasmidial ou elementos genéticos móveis), que confere imunidade adaptativa a estes organismos após uma primeira infecção (Van der Oost *et al.*, 2014). O sistema é constituído pelo *locus* CRISPR, que corresponde a repetições palindrômicas de seqüências curtas que são regularmente espaçadas por seqüências variáveis originadas do DNA invasor-denominadas protoespaçadoras - e dos genes *cas* que codificam endonucleases e se localizam adjacente ao locus CRISPR (Van der Oost *et al.*, 2014). A imunidade adaptativa acontece em três

estágios: i) estágio de aquisição, no qual o DNA invasor entra na célula bacteriana, é fragmentado e em seguida, seleciona-se o fragmento que possui um motivo adjacente ao protoespaçador (motivo PAM) para ser incorporado ao *locus* CRISPR; ii) estágio de expressão, em que o *locus* CRISPR é transcrito completamente, gerando um pré-crRNA que é processado em pequenos crRNAs; e iii) estágio de interferência, momento em que o crRNA processado forma um complexo com a proteína Cas, que é capaz de reconhecer um DNA invasor por complementariedade de bases, e pela ação da Cas esse material invasor é degradado, impedindo a nova infecção (Van der Oost *et al.*, 2014). O reconhecimento e especificidade do alvo por crRNA é possível devido a presença do motivo PAM, que discrimina sequências que fazem ou não parte do genoma da bactéria (Mei *et al* 2016).

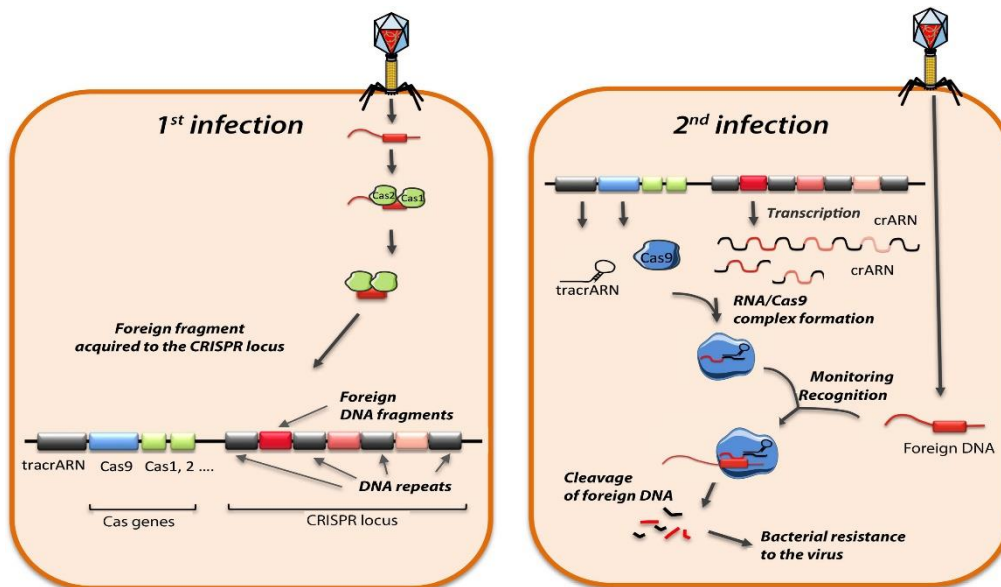


Figura 2: Esquema da imunidade adquirida conferida pelo sistema CRISPR-Cas9 em bactérias após uma infecção viral. Fonte: Duroux-Richard *et al.* (2017).

O sistema CRISPR é altamente diversificado e por isso é dividido em três tipos: I, II e III. Cada tipo apresenta suas características e diferenças, sendo a proteína Cas também diferente entre os tipos (Mei *et al.*, 2016). No tipo II, a endonuclease é a Cas9 e além dos constituintes descritos acima, o sistema possui um gene que codifica um trans-ativador (*tracrRNA*) para facilitar o processamento do pré-crRNA. O *tracrRNA* possui uma região de complementariedade de bases ao pré-crRNA originando um duplex de RNA que recruta a Cas9. Na presença da Cas9, o duplex é clivado pela ribonuclease RNase III formando um híbrido crRNA-*tracrRNA*, que permanece ligado à Cas9 (Van der Oost *et al.*, 2014). O reconhecimento do alvo pelo híbrido é feito por uma porção de 20

nucleotídeos da sequência protoespaçadora presente no crRNA capaz de se ligar ao DNA invasor (Sander *et al.*, 2014). A proteína Cas9 possui um domínio de nuclease em sua estrutura composto pelas nucleases RuvC e HNH, que fazem a quebra na fita dupla do DNA. RuvC age na fita não complementar ao híbrido, enquanto HNH atua na fita complementar (Mei *et al.*, 2016).

Por se tratar de um sistema simples e de alta especificidade, o sistema CRISPR tipo II, principalmente o da bactéria *Streptococcus pyogenes*, foi adaptado biotecnologicamente para se tornar uma ferramenta de edição de genomas, representando um sistema fácil de desenvolver, de alta eficiência, com capacidade de edição de vários genes do mesmo organismo e aplicado a uma variedade de tipos celulares (Ran *et al.*, 2013). O que o sistema sintético aplicado a biotecnologia faz é mimetizar o sistema natural das bactérias, sendo ele constituído pela proteína Cas9 de *S. pyogenes* e por um RNA guia (gRNA), que é a fusão do crRNA e tracrRNA que existe naturalmente nas bactérias (Tsai *et al.*, 2016).

O crRNA sintético corresponde a uma sequência de 20 nucleotídeos complementar ao gene que se deseja editar e que se posiciona adjacente ao motivo PAM no DNA alvo. Para *S. pyogenes*, a sequência PAM reconhecida pela Cas9 é 5' NGG (sendo N qualquer nucleotídeo) e localiza-se na fita não complementar ao gRNA (Mei *et al.*, 2016). A sequência PAM é específica e não pode variar, caso contrário, a Cas9 não reconhece o sítio de clivagem e a quebra da dupla fita do DNA alvo não acontece. A ausência de PAM no gene em que se deseja editar impede a utilização do CRISPR, sendo está uma limitação do sistema (Ran *et al.*, 2013). Outra limitação são os efeitos *off-targets*, ou seja, clivagens que acontecem fora do alvo e comprometem a especificidade do sistema. Para minimizar esse efeito, alguns cuidados devem ser tomados ao desenhar o RNA guia, como: a análise completa do genoma do alvo para avaliar a presença de sequências similares à sequência do RNA guia e o desenho de gRNAs com no máximo três mal pareamentos com o gene alvo, sendo que estes são menos tolerados na extremidade 3' (Tsai *et al.*, 2016; Ran *et al.*, 2013).

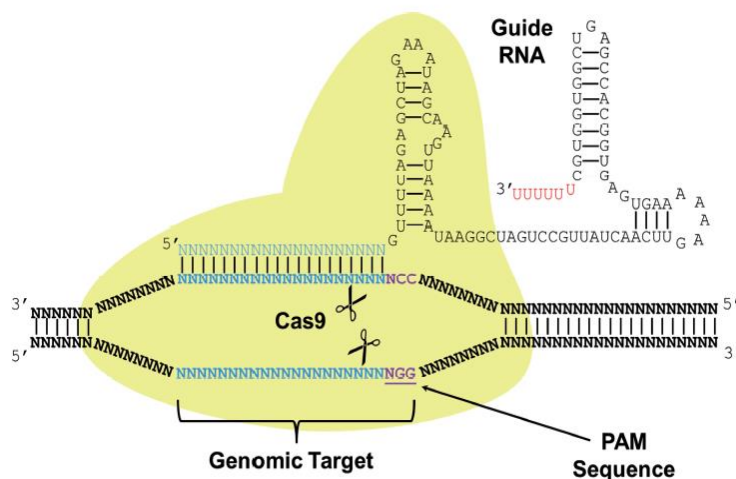


Figura 3: Ligação do RNA guia ao alvo por complementariedade de bases de 20 nucleotídeos adjacente a sequência PAM na fita não complementar. Fonte: DiCarlo *et al.* (2013).

Após a clivagem pela Cas9, os sistemas de reparo das células são ativados para reparar o dano. A quebra de fita dupla pode acionar dois sistemas de reparo diferentes: o sistema de junção de extremidades não homólogas (NHEJ) e o reparo direcionado por homologia (HDR) (Tsai *et al.*, 2015). O NHEJ é um sistema propenso a erros que promove a introdução de mutações por inserções ou deleções de nucleotídeos, e isso pode alterar a fase de leitura dos códons durante a tradução, inserindo por exemplo, um códon de parada prematuro. Pode também alterar o sítio de ligação de fatores de transcrição quando a hidrólise ocorre em regiões promotoras (Sander *et al.*, 2014; Ran *et al.*, 2013). O HDR pode ser usado para introduzir mutações de ponto, inserções e deleções através da recombinação entre o gene alvo e um DNA exógeno com a modificação que se deseja inserir. Assim, para aplicar o CRISPR com o sistema HDR é necessário introduzir um oligonucleotídeo de DNA, de fita simples ou fita dupla, que contenha uma região de complementariedade de bases com o alvo para que haja recombinação (Mei *et al.*, 2016; Ran *et al.*, 2013).

Conhecendo o funcionamento do sistema biotecnológico, várias são as modificações que podem ser feitas para aumentar a aplicabilidade do sistema. Um exemplo é o desenvolvimento de uma Cas9 que possui os domínios de nuclease inativados por mutações específicas, originando a *dead* Cas9 (dCas9). A dCas9 pode funcionar como um bloqueador da transcrição de um gene alvo através do impedimento físico causado pelo complexo Cas9-gRNA, como também pode funcionar como um ativador da

expressão gênica pela ligação de um domínio de ativação em sua estrutura que irá induzir a expressão do gene de interesse (Mei *et al.*, 2016).

O sistema CRISPR-Cas9 já foi aplicado para editar genomas de diversos organismos, como leveduras (DiCarlo *et al.*, 2013), fungos filamentosos (El-Sayed *et al.*, 2017), plantas (Mercx *et al.*, 2017), ratos (Chen *et al.*, 2017) e humanos (Halim *et al.*, 2017).

Lee *et al.* (2016) conseguiram mutantes auxotróficos de uma linhagem industrial de *Saccharomyces cerevisiae* JHS200 utilizando o sistema CRISPR-Cas9, e a partir desses mutantes ele pôde introduzir vetores para expressão heteróloga de enzimas do metabolismo de xilose, observando uma maior produção de etanol a partir de hidrolisados celulósicos. Porém, foi observado o acúmulo de xilitol como subproduto da fermentação de xilose, que surge devido a um desbalanço redox dentro da célula devido a não reoxidação de NADH a NAD<sup>+</sup>. Para resolver esses problemas, os pesquisadores expressaram no mutante auxotrófico a enzima NADH oxidase, resultando em uma maior produção de etanol e menor produção de xilitol.

Para *Kluyveromyces marxianus* existem trabalhos na literatura validando diferentes sistemas CRISPR-Cas9 para edição gênica nesta levedura. Juergens *et al.* (2018), desenvolveram um vetor chamado pUDP002 que possui dois cassetes de expressão constitutiva, um para expressão do gRNA e para expressão da Cas9. A expressão da Cas9 é controlada pelo promotor *TEF1* oriundo de *Arxula adenivorans*. O gRNA é clonado no site de clivagem da *Bsa* I e a inserção é flanqueada por duas ribozimas. A expressão do gRNA é controlada pelo promotor *TDH3* e terminador *CYC1* de *S. cerevisiae*. Ao ser expresso, as ribozimas são clivadas, liberando no núcleo da levedura o gRNA maduro com a região de complementariedade de bases com o *ADE1* (gene alvo envolvido na biossíntese de purinas) e possuindo a estrutura secundária de RNA necessária pelo reconhecimento da Cas9 ao ponto de clivagem. Este sistema foi validado em duas linhagens de *K. marxianus*, a linhagem haploide NBRC1777 e a linhagem diploide CBS397, utilizando tanto o sistema de reparo por homologia quanto o NHEJ. O sucesso da edição foi visto pelo aparecimento de colônias vermelhas nas placas, indicando o acúmulo de pigmento vermelho devido à falta de adenina nas leveduras, além da confirmação por sequenciamento de Sanger.

O CRISPR-Cas9 é uma ferramenta de edição poderosa que tem se destacado pela eficiência de edição e pela simplicidade do sistema. Porém, para *Kluyveromyces marxianus* ainda falta na literatura uma maior aplicabilidade do sistema para estudos de aumento da tolerância ao etanol e aumento da eficiência desta levedura na fermentação

de pentoses. Para a linhagem *K. marxianus* CCT7735 que foi estudada no presente trabalho, não há nenhuma publicação aplicando o sistema.

### 3. OBJETIVOS

#### 3.1 Objetivos Gerais

Avaliar o efeito da deleção de *APJ1* no transcriptoma de *Saccharomyces cerevisiae* sob condição de estresse por etanol e validar o sistema CRISPR- Cas9 para silenciamento de *APJ1* em *Kluyveromyces marxianus*.

#### 3.2 Objetivos Específicos

- Analisar o transcriptoma de *Saccharomyces cerevisiae* BY4741 e *Saccharomyces cerevisiae* Y02999 ( $\Delta apj1$ ) em condição de estresse por etanol através da metodologia de RNAseq, buscando genes diferencialmente expressos;
- Confirmar os dados de expressão para os genes de interesse através de qPCR;
- Desenhar o RNA guia (gRNA) para edição *APJ1* de *K. marxianus* CCT7735  $\Delta ura3$  e clonar no vetor de expressão pML04;
- Validar a expressão de Cas9 em *S. cerevisiae* AH109 e em *K. marxianus* CCT7735  $\Delta ura3$ ;
- Confirmar a edição de *APJ1* em *K. marxianus* CCT7735  $\Delta ura3$ .

## 4. MATERIAL E MÉTODOS

### 4.1 Microrganismos e condições de cultivo

As leveduras utilizadas neste trabalho foram as linhagens de *S. cerevisiae* BY4741 e *S. cerevisiae* Y02999, cedidas pelos professores Ieso de Miranda Castro e Rogelio Lopes Brandão, do Laboratório de Biologia Celular e Molecular da Universidade Federal de Ouro Preto, obtidas da coleção da EUROSCARF; a linhagem de *K. marxianus* CCT7735  $\Delta$ ura3, cedida pelo professor Wendel Batista da Silveira, do Laboratório de Fisiologia de Microrganismos da Universidade Federal de Viçosa; a linhagem de *S. cerevisiae* AH109 (MATa, Trp1-901, leu2-3, 112, ura3-52, his3-200, gal4, LYS2:GAL1UAS-GAL1TATAHIS3, MEL1, GAL2UAS-GALTATA::MELUAS-MEL1TATA-lacZ) pertencente ao Laboratório de Biotecnologia Molecular (LBM) da Universidade Federal de Viçosa.

As culturas de leveduras foram mantidas em meio YPD [1 % (m/v) de extrato de levedura, 2 % (m/v) de peptona, 2 % (m/v) de glicose] acrescido de glicerol 30 % (v/v) a -80 °C e em meio YPD sólido [adição de 1,5 % ágar (m/v)] a 4 °C, quando utilizadas com frequência durante a realização dos experimentos. Somente para a *S. cerevisiae* AH109, o meio YPD foi acrescido com 0,003 % de adenina (YPDA).

As bactérias utilizadas foram as linhagens de *Escherichia coli* DH5 $\alpha$ , pertencente ao LBM e a *E. coli* GM2163 (dam- e dcm-), cedida pela professora Suely Lopes Gomes do Instituto de Química da Universidade de São Paulo. As culturas foram mantidas em meio Luria-Bertani (LB) acrescido de glicerol 30 % (v/v) a -80 °C e em meio LB sólido [adição de 1,5 % ágar (m/v)] a 4 °C, quando utilizadas com frequência durante a realização dos experimentos.

### 4.2 Tolerância a etanol das linhagens *S. cerevisiae* BY4741 e Y02999

A avaliação da tolerância ao etanol foi feita em meio sólido e em leitor de microplaca. Para a avaliação em meio sólido, foram preparadas placas contendo meio YPD (2 % de glicose + 1,5 % de ágar) acrescido de 0 %, 12 %, 16 % e 18 % de etanol (v/v) ultrapuro (Sigma). Pré-inóculos das linhagens *S. cerevisiae* BY4741 e Y02999 foram cultivados em meio YPD (2 % de glicose) por dois dias a 30 °C sob agitação. Após esse tempo, os pré-inóculos foram diluídos em meio YPD (D.O.<sub>600</sub> ~ 0,5) e a partir desta diluição foram feitas mais três diluições sucessivas: 10 x, 100 x e 1000 x. Cinco  $\mu$ L destas diluições foram pipetados em triplicata técnica nas placas contendo as diferentes

concentrações de etanol para as duas linhagens. As placas foram incubadas em estufa a 30 °C e foi feito um acompanhamento visual do crescimento das colônias por 11 dias.

No ensaio em leitor de microplaca foi feita uma curva de crescimento das duas linhagens em meio YPD (2 % de glicose) acrescidos de 0 %, 12 %, 14 %, 16 % e 18 % de etanol (v/v) ultrapuro (Sigma). Pré- inóculos de *S. cerevisiae* BY4741 e Y02999 foram preparados em YPD (2 % de glicose) e cultivados por 16 horas a 30 °C sob agitação. Em seguida, estes pré- inóculos foram diluídos (D.O.<sub>600</sub> ~ 0,15) em um volume final de 200 µL de meio YPD, já acrescidos das diferentes concentrações de etanol. O ensaio foi realizado em sete replicatas técnicas para cada concentração de etanol. A microplaca foi incubada no leitor por 60 horas a 30 °C, com agitação periódica (10 minutos) seguido de mensuração da densidade ótica, permitindo a obtenção de cinéticas de crescimento.

### 4.3 Estratégia experimental- RNAseq

#### 4.3.1 Extração de RNA total de *S. cerevisiae* Y02999 e BY4741 e análise da qualidade

Primeiramente, as linhagens de *S. cerevisiae* Y02999 e BY4741 foram crescidas por 16 horas em meio YPD (2 % glicose) a 30 °C sob agitação, em duplicata biológica, para obtenção de densidade celular. Em seguida, foi feito um novo inóculo (D.O.<sub>600</sub> ~ 1,0) em meio YPD (2 % glicose) adicionando 8 % de etanol (v/v) ultrapuro (Sigma). As células ficaram sob estresse por etanol durante 2 horas a 30 °C sob agitação. Após esse tempo, 3 mL de células foram coletadas por centrifugação (5000 x g, 5 minutos) e lavadas com água ultrapura.

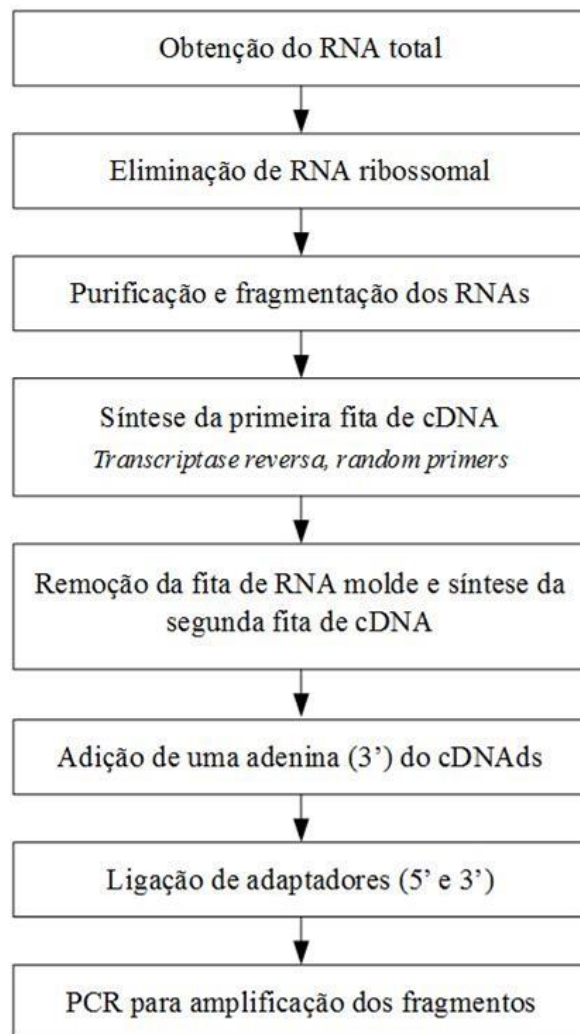
A extração de RNA foi conduzida pelo método *hot phenol* descrito por Collart *et al.* (2001), com modificações. O *pellet* foi ressuspendido em 500 µL de tampão AE (acetato de sódio 50 mM e EDTA 10 mM, pH 5,2) e 50 µL de SDS 10 %. Adicionou-se 550 µL de fenol ácido (pH 3,0) e esta mistura foi incubada a 65 °C em banho seco por 60 minutos, com agitação em vórtex a cada 10 minutos. Ao final, as amostras foram colocadas a 4 °C por 5 minutos, seguido de -20 °C por 20 minutos. Centrifugou-se as amostras (12000 x g, 5 minutos, 4 °C) e a parte aquosa da separação de fases foi transferida para um novo tubo, onde foi adicionado 1 volume (V) de fenol:clorofórmio na proporção 1:1, misturado por inversão e centrifugado (12000 x g, 5 minutos, 4 °C). A fase superior foi transferida para um novo tubo, onde foi adicionado 1 V de clorofórmio, aguardado 5 minutos, centrifugado (12000 x g, 5 minutos, 4 °C) e a fase superior transferida novamente. Esta etapa foi repetida mais uma vez e a fase superior transferida para um novo tubo, onde foi

adicionado 0,1 V de acetato de sódio 3 M, pH 5,3 e 2,5 V de etanol absoluto gelado. A precipitação ocorreu por 16 horas a -20 °C e centrifugou-se (12000  $\times$  g, 5 minutos, 4 °C) para obtenção do *pellet*, que foi lavado uma vez com 500  $\mu$ L de etanol 80 % e novamente centrifugado (12000  $\times$  g, 5 minutos, 4 °C). Por fim, o *pellet* foi solubilizado em 30  $\mu$ L de água ultrapura.

A qualidade do RNA extraído foi avaliada em gel de agarose 1,5 % corado com brometo de etídeo 0,1  $\mu$ g/mL e submetido a corrida eletroforética em tampão TAE 1 X (Tris Acetato EDTA). A quantificação do RNA extraído foi feita por espectrofluorometria utilizando o aparelho Qubit 3.0 (Invitrogen), opção RNA *Broad Range*, seguindo as instruções do fabricante. Cinco  $\mu$ g de RNA foram tratados com uma unidade de DNase I, *Amplification Grade* (Invitrogen), para eliminação de possíveis contaminações com DNA e quantificado novamente no Qubit. Uma segunda análise da qualidade dos RNAs tratados foi realizada utilizando o sistema *Agilent 2100 Bioanalyzer*, seguindo o guia *Agilent RNA 6000 Nano Kit Quick Start Guide*.

#### 4.3.2 Construção, validação e sequenciamento das bibliotecas

A construção das bibliotecas de DNA complementar dupla fita (cDNAs) a partir do RNA total tratado foi feita utilizando o guia *Illumina Truseq strand total RNA sample preparation Guide (Low sample protocol)*, partindo de 1  $\mu$ g de RNA total. As etapas realizadas pelo guia da Illumina para construção das bibliotecas estão esquematizadas no fluxograma 1.



Fluxograma 1: Etapas para a construção de bibliotecas de cDNAs.

Para validação das bibliotecas, foi feita uma análise quantitativa e qualitativa dos cDNAs. A quantificação das bibliotecas foi realizada através de qPCR, com primers que se anelam em uma região consenso dos adaptadores, seguindo o *Kapa Library Quantification Kit (Illumina Platforms)* e utilizando o aparelho Step One (*Applied Biosystems*). Foi construída uma curva-padrão utilizando os seis padrões de concentrações conhecidas do *Kit* (20, 2, 0,2, 0,02, 0,002 e 0,0002 pM), relacionando o *threshold cycle* (Ct) médio da duplicata técnica com o  $\log_{10}$  da concentração dos padrões. Para as bibliotecas foram feitas diluições seriadas de 5.000 x, 50.000 x e 500.000 x e a partir dos valores de Ct médio da duplicata técnica de cada diluição foi possível calcular a concentração de cada biblioteca pela curva-padrão, acertando este valor de acordo com o ajuste da concentração estabelecido no *kit*. Para análise do tamanho médio dos

fragmentos da biblioteca, foi usado o sistema *Agilent 2100 Bioanalyzer* seguindo o guia *Agilent High Sensitivity DNA Kit Quick Start Guide*.

As bibliotecas validadas foram diluídas sucessivamente para concentrações de 50 nM, 20 nM e 2 nM em Tris-HCl 10 mM acrescido de 0,1% de Tween 20. Um *pool* de bibliotecas de 100 µL foi feito colocando 2 µL de cada uma das quatro bibliotecas de levedura a 2 nM. Este *pool* foi desnaturado e diluído para a concentração de 13 pM de acordo com o protocolo *Denature and Dilute Libraries Guide for Hiseq and GAIIX Systems* e injetado no sequenciador Hiseq 2500 do Núcleo de Análise de Biomoléculas da UFV, seguindo os protocolos de preparação do aparelho *Truseq Rapid SBS Kit (200 cycles)* e *Hiseq 2500 System User Guide (Chapter 4 Perform a Rapid Run)*, selecionando o tipo de corrida *paired-end*.

#### 4.3.3 Tratamento dos dados e análises de bioinformática

Após o sequenciamento, os dados brutos foram tratados e analisados por programas de bioinformática. Para avaliar a qualidade das *reads* geradas no sequenciamento foi usado o programa FastQC (versão 0.11.5, Andrews, 2010). Este programa indica vários parâmetros das *reads* sequenciadas, como a quantidade total de sequências, o comprimento das sequências, a % GC das sequências, pontuação de qualidade de todas as bases, pontuação de qualidade das sequências, o conteúdo de adaptadores, o conteúdo N (qualquer nucleotídeo) entre as bases, dentre outros. A eliminação das sequências baseado na análise da qualidade foi feita através do programa Trimmomatic, segundo Bolger *et al.* (2014), eliminando as sequências com valor de Phred < 15. Também foram eliminadas as 10 bases do começo e do final de cada *read*, assim como as *reads* com comprimento abaixo de 25 pares de bases. Após a trimagem, foi feita uma outra análise no FastQC para ver se as sequências de baixa qualidade tinham sido eliminadas corretamente.

A montagem do transcriptoma foi feita pelo alinhamento das *reads* ao genoma de referência (*Saccharomyces cerevisiae* S288c). Foi utilizado o programa Bowtie2 (versão 2.2.8, Langmead *et al.*, 2012), para indexar o genoma de referência e o programa TopHat (versão 2.1.1, Trapnell, 2009), para fazer o mapeamento ao genoma indexado. A montagem dos transcritos e a contagem das *reads* mapeadas foi feita utilizando o programa Cufflinks (Trapnell *et al.*, 2010), sendo o resultado expresso em fragmentos por kilobase de transcrito por milhão de fragmentos mapeados (FPKM). O FPKM é um valor normalizado da contagem das *reads* mapeadas que considera o tamanho do transcrito. O

cálculo de *Fold Change* (FC) foi feito dividindo o valor de FPKM dos genes da linhagem *S. cerevisiae* Y02999 pelo FPKM da linhagem *S. cerevisiae* BY4741. Para linearização dos valores de diferença de expressão foi calculado o  $\log_2$  FC.

#### 4.3.4 Validação da expressão de alguns genes por qPCR

Quatro genes (*ARG3*, *BTN2*, *COS12* e *GMC2*) que apresentaram diferenças no valor de FPKM entre as duas linhagens na análise pelo RNAseq e quatro genes (*OLE1*, *HSP104*, *UBC9* e *SIZ1*) que na literatura apresentam alguma relação ao processo de tolerância ao etanol, tiveram sua expressão analisada por qPCR. A extração de RNA total em triplicata biológica, análise da qualidade e quantificação foram feitas de acordo com o item 4.3.1. Além da extração de RNA total com o estresse por etanol de 8 % (v/v), foram feitas duas concentrações a mais para este experimento: 0 % (v/v) e 12 % (v/v) de etanol. Para a condição 0 %, foi feito um inóculo de (D.O.<sub>600</sub> ~ 0,8) a partir do pré-inóculo crescido por 16 horas e este inóculo foi incubado a 30°C sob agitação por 2 horas. Para o estresse com 12 % de etanol, a estratégia experimental foi igual a condição 8 % descrita no item 4.3.1. Após a extração, 1µg de RNA total foi tratado com uma unidade de DNase I (Sigma) seguindo as recomendações do fabricante. Foi feita uma PCR com *primer* do gene *IPP1* (95 °C por 5 min; 30 ciclos de 95 °C por 1 min, 60 °C por 30 s e 72 °C por 30 s; 72 °C por 5 min) para confirmar o funcionamento da enzima na eliminação de qualquer DNA contaminante nos RNAs.

A sequência da região codificadora dos oito genes diferencialmente expressos e de mais dois genes de expressão constitutiva em condições de estresse por etanol, *ACT1* (Syahidah *et al.*, 2013) e *RBS1* (Ma *et al.*, 2010), foram obtidas no banco de dados *Saccharomyces cerevisiae* Genome Database. O desenho dos *primers* foi realizado na plataforma online *GenScript Real-Time PCR (TaqMan) Primer Design*. Para garantir a especificidade dos *primers* no gene alvo, foi feita uma análise de alinhamento no genoma completo de *Saccharomyces cerevisiae* S288c utilizando o programa e-PCR. Foi feita uma PCR com todos os 10 *primers* (ciclo descrito neste item) para confirmar o pareamento destes no genoma de *S. cerevisiae* BY4741 e Y02999 a 60°C, temperatura na qual o experimento de qPCR foi conduzido.

Tabela 1: Oligonucleotídeos específicos para os genes envolvidos na análise de qPCR.

<b>Genes <i>S. cerevisiae</i></b>	<b>Primer F (5'-3')</b>	<b>Primer R (5'-3')</b>
<i>ARG3</i>	AGGCCTCCACCAATGCCAAT	TGGATCAGCCACAGAGACAAGT
<i>BTN2</i>	CGAAAGAACCCTAACCACAAAGAAGACC	CCTTCAATCTGGCTTCCTTCTCTTTC
<i>COS12</i>	CGTTACTGCTCTTATGTTTGAAGTGCT	ACACGTTCCCAGTCTTCAGTGT
<i>GMC2</i>	CCGCCGAATGGCAAATCAACA	GTTGTTTGGACATCATCCCATCTTGAC
<i>OLE1</i>	ACTGGCACCAACATTTGAACTGG	AACACCACCGACAGCGTAGT
<i>HSP104</i>	CTGAGCAGGCTCGTCAAGGT	CGGTCTTACCGATACTGGCT
<i>UBC9</i>	ACGTGTATCCAAGTGGCACAATATG	CCATGCAGGCTCTTGAGCAG
<i>SIZ1</i>	TCAAGAGTAAGCCGGGCACA	CTTGTATTCTTCGTGGTGAAGGC
<i>ACT1</i>	TGGATTCCGGTGATGGTGT	TCAAAATGGCGTGAGGTAGAGA
<i>RBS1</i>	GGTGCGATTCCGAAACGTAGTG	CAGTATCCCTAGATGACGGCGAA
<i>IPPI</i>	TACTTCCCAGGTCTGTTGAGG	AACAGAACCGGAGATGAAG
<i>APJ1</i>	CATGCAAAGGGCAAGGTATT	CGCAATCTGCACAAGTTTGA

A síntese de cDNA foi realizada com o *High Capacity cDNA Reverse Transcription Kit (Applied Biosystems)* utilizando 1µg de RNA total, seguindo as especificações do fabricante. Para confirmar a síntese do cDNA foi feita uma PCR com os *primers* do gene *IPPI* (ciclo descrito no item 4.3.4). Também foi feita uma PCR com os *primers* do gene *APJ1* (ciclo descrito no item 4.3.4) para confirmar a deleção deste gene nos cDNAs da linhagem *S. cerevisiae* Y02999 e mostrar a presença nos cDNAs da linhagem *S. cerevisiae* BY4741.

Os ensaios de qPCR foram realizados utilizando o *Power SYBR Green PCR Master Mix (Applied Biosystems)* na concentração final 1 x (seguindo o manual do fabricante) e conduzidos no aparelho *Step One (Applied Biosystems)*. Para todos os 10 genes foi construída uma curva-padrão relacionando o Ct médio da duplicata técnica com o log<sub>10</sub> da concentração de cDNA, a partir de um *pool* de cDNAs de *S. cerevisiae* BY4741 e Y02999 diluído para as concentrações de 50, 10, 2, 0,4, 0,08, 0,016 e 0,0032 ng/µL, e utilizando uma concentração final de 1,0 µM de *primer* F e *primer* R. A montagem das placas de expressão foi feita para os 10 genes em triplicata biológica dos cDNAs das duas linhagens nas três concentrações de etanol, diluindo 10 vezes os cDNAs que estavam a 50 ng/µL e mantendo concentração final de 1,0 µM para os primers. Incluiu-se uma duplicata técnica no experimento.

A quantificação da expressão gênica foi realizada usando o método de quantificação absoluta por curva-padrão, tendo os genes *ACT1* e *RBS1* como normalizadores dos dados. Primeiramente, foi feita a média dos valores de Cts da duplicata técnica e pela substituição destes valores nas curvas-padrões construídas para cada gene, foi encontrado a quantidade de mRNA em ng/μL para cada um dos 10 genes nas três concentrações de etanol e em triplicata biológica. Os dados foram normalizados dividindo as concentrações dos oito genes pelas concentrações dos dois genes de expressão constitutiva (*ACT1* e *RBS1*) e em seguida, foi feita a média dos valores encontrados nas divisões. As análises estatísticas foram conduzidas no *GraphPad Prism 5* utilizando Anova e o pós teste de comparação múltipla Bonferroni para comparação das médias encontradas para cada concentração de etanol nas duas linhagens.

#### 4.4 Estratégia experimental- sistema CRISPR-Cas9

##### 4.4.1 Desenho do gRNA e construção do vetor pML104-gRNA

Primeiramente, buscou-se a sequência de *APJ1* de *K. marxianus* DMKU3-1042 no NCBI. Através de um programa desenvolvido *in house* pelo professor Dr. Tiago Antônio de Oliveira Mendes, foi desenhado um gRNA de 20 nucleotídeos vizinho a uma sequência 5' NGG (sendo N qualquer nucleotídeo) do gene *APJ1*. Dentre as possibilidades de gRNA geradas pelo programa, foi escolhido aquele que ao realizar o BLAST não anelava em nenhuma outra região do genoma de *K. marxianus* e era próximo ao ATG inicial de tradução de *Apj1p*.

Para aplicação do sistema CRISPR-Cas9 em *K. marxianus* foi escolhido o vetor pML104 desenvolvido por Laughery *et al.* (2015), figura 4. Este vetor tem a expressão da endonuclease Cas9 controlada pelo promotor constitutivo pTDH3 de *S. cerevisiae*, possui um cassete para expressão do gRNA controlado pelo promotor pSNR52 de *S. cerevisiae*, bem como marca de seleção para transformação em bactérias (AmpR) e em leveduras que possuem mutação auxotrófica no gene *URA3*.

A clonagem do gRNA dentro do cassete de expressão acontece por meio de duas enzimas de restrição, *Bcl I* e *Swa I*. Ao clivar o vetor, essas enzimas geram, respectivamente, uma extremidade coesiva e uma extremidade cega. Sendo assim, para completar o desenho do oligonucleotídeo escolhido como gRNA (oligo 1) adicionaram-se uma extensão na extremidade 5' complementar à quebra gerada pela *Bcl I* (sequência em vermelho no oligo1 mostrado na tabela 2) e 13 nucleotídeos na extremidade 3'

(Laughery *et al.*, 2015) correspondente à parte do gRNA necessária para a atividade de Cas9 (sequência em verde no oligo1 mostrado na tabela 2). Por fim, desenhou-se um segundo oligonucleotídeo (oligo 2) complementar ao primeiro, deixando somente a extensão 5' livre.

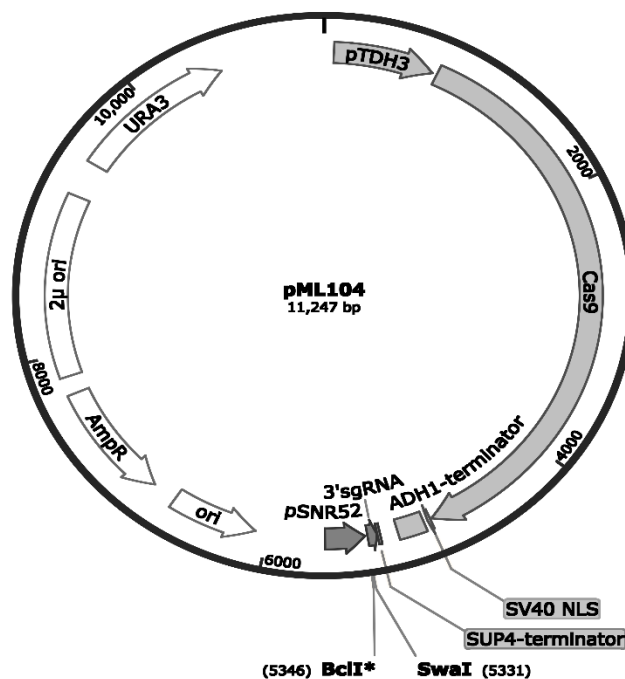


Figura 4: Vetor pML104. Fonte: Laughery *et al.* (2015).

Para clonagem do gRNA, primeiramente uma colônia de *Escherichia coli* DH5 $\alpha$  contendo o vetor pML104 foi crescida por 16 horas, a 37 °C sob agitação em meio LB contendo 100  $\mu$ g/mL de ampicilina. Em seguida foi feita a extração do vetor utilizando o *Easy Plasmid MiniPrep Kit (Trans)*, seguindo as recomendações do fabricante. Quatro  $\mu$ g de vetor foi digerido simultaneamente com as enzimas *Swa I (Thermo Scientific)* e *Ksp22 I (SibEnzyme)*. A reação de digestão consistiu de 4  $\mu$ g do vetor, 2 unidades da enzima *Swa I*, 4 unidades da enzima *Ksp22 I*, 2  $\mu$ L de tampão Orange 10 x (*Thermo Scientific*) e água para completar o volume de 20  $\mu$ L. A reação aconteceu por uma hora a 30 °C e a inativação das enzimas ocorreu por 20 minutos a 65 °C, ambas no termociclador. A análise da digestão foi feita por eletroforese em gel de agarose 0,8 %, sendo que a banda correspondente ao plasmídeo linear foi excisada e purificada utilizando o kit *Wizard SV Gel and PCR Clean-Up System (Promega)*, seguindo as recomendações do fabricante.

O oligo 1 e o oligo 2 desenhados como gRNA foram fosforilados na extremidade 5' pela T4 Polinucleotídeo Kinase (T4 PNK, *Thermo Scientific*) seguindo as indicações

do fabricante. A reação consistiu de 2  $\mu\text{L}$  de oligo (1 ou 2) a 100  $\mu\text{M}$ , 2  $\mu\text{L}$  de tampão A 10 x, 2  $\mu\text{L}$  de dATP a 10mM, 1  $\mu\text{L}$  (10 unidades) de T4 PNK e 13  $\mu\text{L}$  de água ultrapura. A reação foi incubada a 37 °C por 20 minutos e a enzima inativada pelo calor a 75 °C por 10 minutos, ambas no termociclador.

A hibridização dos oligos fosforilados foi feita seguindo Laughery *et al.* (2015), com modificações. 1  $\mu\text{L}$  de cada oligo a 10  $\mu\text{M}$  foram colocados em 1  $\mu\text{L}$  do tampão da T4 ligase 10 x (Promega) mais 3,7  $\mu\text{L}$  de água ultrapura. A hibridização ocorreu no termociclador, onde inicialmente as amostras ficaram incubadas a 95 °C por 6 minutos e depois houve um decréscimo de 1 °C/ minuto por 70 vezes até a temperatura de 25 °C.

Para reação de ligação do vetor linear ao híbrido, uma unidade de T4 ligase (Promega) e aproximadamente 100 ng de vetor linear purificado foram adicionados à reação de hibridização. A reação de ligação foi incubada a 37 °C por uma hora seguido de *overnight* a 4 °C. Após, 4  $\mu\text{L}$  da reação de ligação foi utilizado para transformação de *E.coli* DH5a pela metodologia de choque térmico descrita por Sambrook *et al.* (1989) e a seleção dos transformantes foi feita em meio LB sólido contendo 100  $\mu\text{g/mL}$  de ampicilina. O vetor clonado com o guia foi chamado de pML104-gRNA.

Algumas colônias transformantes foram selecionadas para o PCR de colônia (95 °C por 10 min; 35 ciclos de 95 °C por 1 min, temperatura de anelamento dos *primers* (Ta) por 40 s e 72 °C por 1 min; 72 °C por 7 min) com *primers* do vetor que anelam na região promotora e terminadora do cassette de expressão do gRNA (*primers* seqF e T3R, Ta = 46 °C). Foi feita a extração dos plasmídeos das colônias positivas para o PCR de colônia, utilizando o *Easy Plasmid MiniPrep Kit* (Trans). Um novo PCR combinando o *primer* T3R com o oligo 2 (Ta = 51 °C) foi realizado na tentativa de confirmar o inserto.

Através de PCR com os *primers* T3R e seqF foi amplificado a região de expressão do gRNA de algumas colônias, os fragmentos foram purificados utilizando o kit *Wizard SV Gel and PCR Clean-Up System* (Promega) e preparados para sequenciamento Sanger na Myleus, seguindo as instruções da empresa.

Tabela 2: Oligo 1 e 2 que formam o gRNA e *primers* específicos para identificação de possíveis vetores clonados

<i>Primers</i>	Sequência 5' - 3'
Oligo 1	GATCTACGAAACTTTACGATATTCGTTTTAGAGCTAG
Oligo 2	CTAGCTCTAAAACGAATATCGTAAAGTTTCGTA
seqF	CATGCTTATAACTAATTACATGAC
T3R	GCAATTAACCCTCACTAAAGG

#### 4.4.2 Transformação de *K. marxianus* CCT7735 $\Delta$ *ura3* e *S. cerevisiae* AH109

A transformação de *K. marxianus* CCT7735  $\Delta$ *ura3* foi feita segundo Belo (2013), com modificações. Primeiramente, foi feito um pré- inóculo de *K. marxianus* CCT7735  $\Delta$ *ura3* em meio YPD crescido a 30 °C por 16 horas em agitação. Em seguida, foi feito um novo inóculo em 25 mL de meio YPD (D.O.<sub>600</sub> ~ 0,2). Este inóculo foi crescido a 30°C sob agitação até atingir D.O.<sub>600</sub> igual a 1. Centrifugou-se essa cultura a 4500  $\times$  g por 3 minutos e o *pellet* foi lavado uma vez com água ultrapura estéril. O *pellet* limpo foi ressuspendido em 1,2 mL de Polietilenoglicol (PEG) 50 %, 300  $\mu$ L de acetato de lítio 1M e 15  $\mu$ L de Ditionitroto (DTT) 1M e centrifugado a 2500  $\times$  g por 2 minutos, descartando o sobrenadante. O *pellet* foi ressuspendido em 400  $\mu$ L de PEG 50 % e 100  $\mu$ L de acetato de lítio 1M. Após, 200  $\mu$ L dessa mistura foi transferido para um novo tubo onde foram adicionados 7  $\mu$ L de DNA fita simples (ssDNA) de esperma de salmão a 10 mg/mL (previamente fervido) e 10  $\mu$ g de pML104-gRNA. O choque térmico nas leveduras foi conduzido em banho-maria a 47 °C por 13 minutos. Após esse tempo, as células foram centrifugadas (2500  $\times$  g por 2 minutos) e o *pellet* lavado uma vez com água ultrapura estéril. O *pellet* foi ressuspendido em 900  $\mu$ L de meio YPD (2% de glicose) e deixado em recuperação por uma hora a 30 °C sob agitação lenta. 200  $\mu$ L de células foram plaqueadas em meio mínimo para leveduras sem adição de uracila (SD – URA). As placas foram incubadas a 30°C por um período de 2-3 dias até o aparecimento das colônias.

Para a transformação de *S. cerevisiae* AH109, foi utilizado o protocolo de Gietz *et al.* (2007), com modificações. Uma colônia da levedura foi crescida a 30 °C por 16 horas sob agitação em meio YPDA. Um inóculo de 25 mL foi feito a partir do pré-inóculo em meio YPDA, iniciando em D.O.<sub>600</sub> ~ 0,2 e crescido a 30 °C em agitação até atingir D.O.<sub>600</sub> ~ 0,6. As células então foram coletadas por centrifugação (3.000  $\times$  g por 10 minutos) e lavadas uma vez com água ultrapura estéril. Ao *pellet* foram adicionados 2,5 mL de PEG

50%, 400  $\mu$ L de acetato de lítio 1 M, 300  $\mu$ L de ssDNA de esperma de salmão a 10 mg/mL (previamente fervido) e 10  $\mu$ g de pML104. Essa mistura foi mantida em banho-maria à 42 °C por 50 minutos, sendo homogeneizada a cada 10 minutos. Após este tempo, o *pellet* foi coletado por centrifugação (3.000  $\times$  g por 5 minutos) e lavado duas vezes com água ultrapura estéril para remoção do PEG contaminante. Por fim, as células foram ressuspendidas, plaqueadas em meio SD – URA e incubadas a 30 °C por três dias para o aparecimento das colônias transformantes.

#### 4.4.3 Validação da expressão de Cas9 em *K. marxianus* CCT7735 $\Delta$ ura3 e *S. cerevisiae* AH109

Para validação da expressão de Cas9 foi feita a extração de RNA de *K. marxianus* CCT7735  $\Delta$ ura3 e de *S. cerevisiae* AH109 transformadas com o vetor pML104. Como controle do experimento, também foi extraído RNA das duas linhagens não transformadas. Todas as extrações foram realizadas em triplicata biológica.

As extrações foram conduzidas pelo método *hot phenol* descrito por Collart *et al.* (2001), com modificações (protocolo detalhado no item 4.3.1). Os RNAs foram quantificados no Qubit, opção RNA *Broad Range*, seguindo as instruções do fabricante. Dois  $\mu$ g de RNA foram tratados com uma unidade de DNase I, (Sigma) e a eliminação dos DNAs contaminantes foi confirmada por PCR (95 °C por 5 min; 30 ciclos de 95 °C por 1 min, 60 °C por 30 s e 72 °C por 30 s; 72 °C por 5 min) utilizando *primers* do gene actina (*ACT1*) tanto para *K. marxianus* quanto para *S. cerevisiae*.

A síntese de cDNA foi realizada com o *High Capacity cDNA Reverse Transcription Kit* (*Applied Biosystems*) utilizando os 2  $\mu$ g de RNA tratados, seguindo as especificações do fabricante. Para confirmar a síntese dos cDNAs foi feita uma PCR com os *primers* do gene *ACT1* (ciclo descrito acima).

Para validação da expressão de Cas9, foi desenhado dois pares de *primers* para o gene expresso no vetor utilizando a plataforma online *GenScript Real-Time PCR (TaqMan) Primer Design*. Para garantir a especificidade dos *primers* no gene alvo, foi feita uma análise de alinhamento no genoma completo de *S. cerevisiae* S288c e *K. marxianus* DMKU3-1042 utilizando o programa e-PCR. O anelamento dos *primers* a 60 °C no vetor foi confirmado por PCR (ciclo descrito acima) e todos os dois pares foram testados no genoma das duas leveduras para ver possíveis pareamentos inespecíficos. O protocolo de extração de DNA genômico está descrito no item 4.4.4.

O experimento de expressão foi conduzido por PCR convencional, objetivando observar somente presença ou ausência de expressão. Assim, foi feita uma PCR (ciclo descrito acima) com todos os cDNAs sintetizados das duas leveduras tanto transformadas como não transformadas e o resultado foi visto em gel de agarose 2 % submetido a corrida eletroforética.

Tabela 3: Oligonucleotídeos utilizados para confirmar a síntese de cDNA e a expressão de Cas9

<b>Genes</b>	<b>Primer F (5'-3')</b>	<b>Primer R (5'-3')</b>
<i>ACT1 S. cerevisiae</i>	TGGATTCCGGTGATGGTGTT	TCAAAATGGCGTGAGGTAGAGA
<i>ACT1 K. marxianus</i>	ACGTTGTTCCAATCTACGCC	CTTGTTCGAAGTCCAAAGCG
Cas9 par 1	TGGCTGGTTCACCCGCTATT	TTCATCCGCTCCCTGCTGTT
Cas9 par 2	TGGTGGAACCGCCCTAATC	GGCCTTCCGATCTCCTGCT

#### 4.4.4 Busca por transformantes positivos $\Delta apj1$

Após a transformação de *K. marxianus* CCT7735  $\Delta ura3$  com o pML104-gRNA, foi feita uma seleção aleatória de colônias transformantes, na tentativa de identificar colônias positivas para o funcionamento do sistema CRISPR-Cas9, ou seja, colônias que sofreram alteração na sequência de nucleotídeos de *APJ1*.

As colônias selecionadas foram inoculadas em meio seletivo SD - URA e crescidas por 16 horas sob agitação, para extração de DNA genômico. As células foram coletadas por centrifugação (8000 rpm por 10 minutos) e lavadas uma vez com água ultrapura estéril. O *pellet* foi ressuscitado em 400  $\mu$ L de tampão de extração [2% Triton 100 x (v/v), 1 % SDS (m/v), 0,1 M NaCl (m/v), 0,1 M Tris pH 8,0 (m/v) e 0,001 M EDTA pH 8,0 (m/v)] e transferido para tubos de 1,5 mL. Foram adicionados 0,3 g de *beads* de vidro (Sigma, 425-600  $\mu$ m) a cada tubo e essa mistura foi agitada por 2 minutos no vórtex para lisar as células. Para eliminação das *beads* foi feita uma centrifugação (12.000 x g por 10 minutos) e o sobrenadante foi coletado e transferido para um novo tubo. O DNA foi precipitado com 0,1 V de acetato de potássio (5 M) e 2 V de etanol absoluto gelado. A precipitação ocorreu à temperatura ambiente por 2 minutos. Em seguida, os tubos foram centrifugados (12.000 x g por 5 minutos) e o *pellets* lavados com 500  $\mu$ L de etanol 70 %. Após seco à temperatura ambiente, o *pellet* foi ressuscitado em água ultrapura.

Os DNAs genômicos foram usados em uma PCR (95 °C por 5 min; 35 ciclos de 95 °C por 1 min, 50 °C por 40 s e 72 °C por 1 min; 72 °C por 5 min) em que parte do gene

alvo, onde possivelmente ocorreu a edição, foi amplificado. Para isso foi desenhado um par de *primer* que flanqueia a região onde a edição é esperada. A figura 5 é referente a sequência de nucleotídeos de *APJI* e mostra os pontos de anelamento dos *primers* F e R no gene.

```
5' AACAGAATTACGAGAGTAAAAGAAGTATGGTGAAAGATACGAAACTTTACGATATTCTGGAGGTGAGC
GCGGATGCCACTGATCAGGAGATCAAAAAAGCGTACAGGAAAAAGGCGTTGCGGCACCATCCAGACAAAA
ACAACCATAGCCCAGAGTCTATAAAGTTATTCAGGAGTTAAGTCATGCGTATGAAATATTAAGTGATGA
AAGTAAGAGGGAGTTGTACGATCAGTATGGGACTGTGGATGAGACTGAAATCAATGAGATGATGTCTAAG
CAAAGAGGTGCAGCTAGCAATGCGGCGTTTGGCCACACAGCAGGGGACCTTTTTTGCAGTTTTTTGGCG
GCCGTCCGGGCTCTGCTTCTGGGGCGGGCGCTAACGGTTCTGGCGGGAACAGCTTCTCCAATCGTTTTTC
TCCAGACTTTTCTAGTTATGACATGGATGGGAGCCAGGAGATGGCTTCTGGCCCAGGCATTCACCACAAC
TTGAAGTGACGTTGTACGACTTGTACCACGGGAAGCGTGCCAAGTTAGCACTAAATAGAACTAGGTTGT
GGAAAAATGCCAAGGTTACGGTGGTAAAAAGACATCGCAATGTAAAGGTTGTCAGGGTACGGGTATGTAC
ACTACAACCAAAAAGAATGGGGCCCATGGTTCAAACATGGCAAACCACATGCAAAGATTGTGGAGGTTCTG
GTAAGTTCATTAGAAGTAAGGACGCATGTGAAGAATGTTCTGGAAACGGGTTTCGTCAAGGAACGGAAGAT
ATTTGATGTTGAGATACAGCCTGGTATGACTGATGGTAAAGAGATCGTGTACCAGGTTGAGGCTGATGAA
GTTATAAATACAGAAATATGGTAAGGAACGCGTACACCCAGGTGATGTCATCATCACCATCGAGCTCACTA
GGAGAGATGAACACAACATGAAATATAAGTACATGGTACATGGACATGATCTCATCCTAGACAACCTTTGA
GGTGCAGCTTAAAGACAAGTCTTTGTGGTGGTACCATCGTCATTGAGGACCATCCCAGCGGTAAGCCATTC
AAGATTGATGTGCTTGCAGGTGAGCTCTTGAAACCCGGCTGTATTAAGTGCGTGGAAAACAAAGGTATGC
CTGTCGATCAGTATGGCAAGTTCGGAAACTTGTACATCAGATTCAGGTTAAATTTCCAGAGCAGCTAAA
AAACGAAACCATTTGAAAAGTTATCGCAAGCTCTTGTGGAAGATGAGAACATAATAAAGCAACCGCCAAAT
GATGACAAGAACGGCGGATTGATCCTTACAGGTCATGGGAAGAACAAGTGCTCAGTAGCTTTACTTGCTG
ATGATTTACAGTCGAAGTTTAGAGAAGCAAAGAAAAACCCTTCAAATAAGAGGAAACGGGGAGAACATGG
TGCATACAATAAAAAATGGCACGCAATTCGATAGCAGTTATGATCCATCAGATTCCTGCCAGGTGAATTGA
AATGAAAAGTTGTAACATATATAGAAAAGAAAATCAAAAATCCACTTTCAATTTCC 3'
```

Figura 5: Sequência de nucleotídeos de *APJI K. marxianus* DMKU3-1042. As trincas ATG e TGA em vermelho representam os códons de início e final da tradução, respectivamente. O TGG em verde corresponde à sequência 5'-NGG escolhida para guiar a edição do gene. Os dois conjuntos de sequências sublinhados correspondem aos *primers* F e R desenhados para amplificação de parte do alvo e confirmação da edição pelo sistema CRISPR/Cas9. O gRNA se ligará por complementariedade de bases à fita 3'-5' em uma posição vizinha ao NGG presente na fita oposta e a clivagem promovida pela Cas9 possivelmente ocorrerá na região entre o ATG inicial e o NGG na fita 5'-3'.

O fragmento correspondente ao tamanho do amplicon esperado foi excisado do gel de agarose 1 % e purificado utilizando o kit *Wizard SV Gel and PCR Clean-Up System* (Promega). O fragmento devidamente preparado foi enviado para sequenciamento de Sanger na empresa Myleus.

## 5. RESULTADOS

### 5.1 Ensaio de tolerância ao etanol comparando *S. cerevisiae* Y02999 e BY4741

Após 11 dias de incubação, o ensaio de tolerância ao etanol em meio sólido mostrou que as linhagens *S. cerevisiae* Y02999 e BY4741 cresceram de forma semelhante em meio com 0 % e 12 % (v/v) de etanol, porém em 16 % (v/v) a linhagem Y02999 cresceu melhor que a linhagem BY4741 em todas as diluições plaqueadas. Em 18 % (v/v), a BY4741 não cresceu em nenhuma diluição enquanto Y02999 apresentou ligeiro crescimento nas duas primeiras diluições (D.O. 600 0,5 e 0,05). Os resultados deste ensaio são mostrados na figura 6.

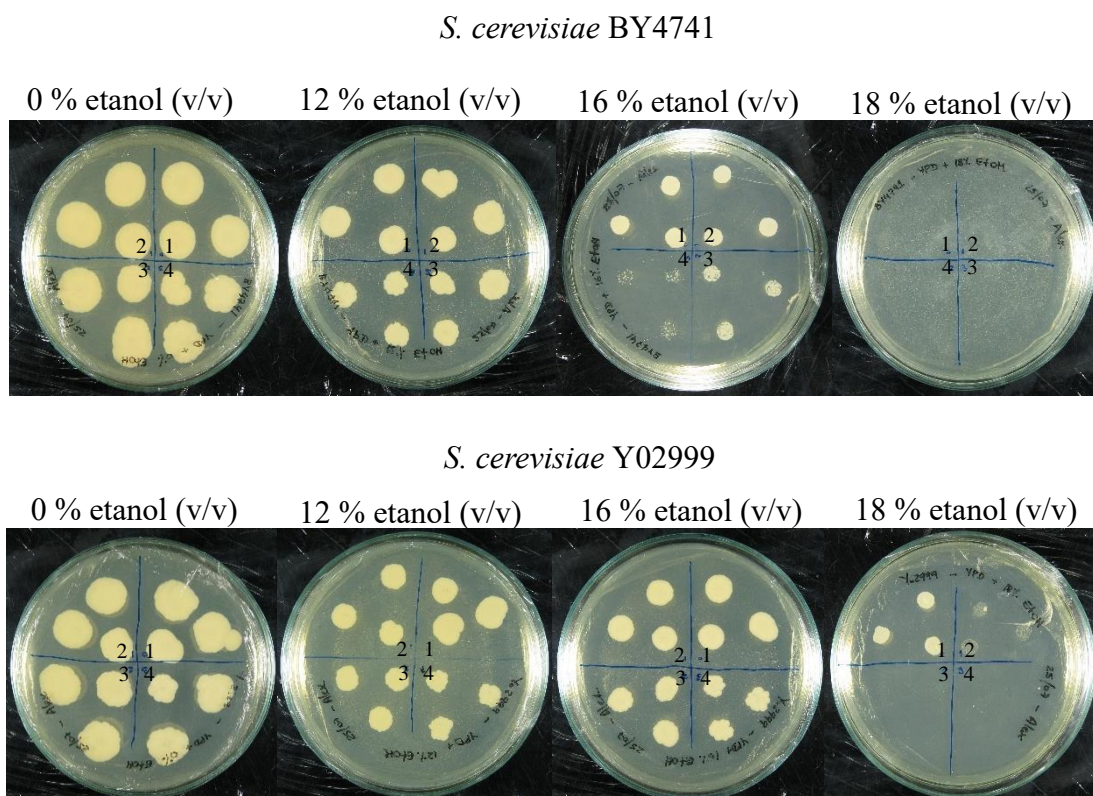


Figura 6: Ensaio de tolerância ao etanol em meio sólido. Crescimento de *S. cerevisiae* BY4741 e Y02999 em meio YPD com diferentes concentrações de etanol [0 %, 12 %, 16 % e 18 % (v/v)] após 11 dias de incubação a 30 °C. O ensaio foi realizado em triplicata técnica para cada diluição plaqueada (D.O. 600 0,5, 0,05, 0,005 e 0,0005), sendo que cada diluição está em um quadrante da placa. Os números 1-4 escritos no centro da placa representam os quadrantes, sendo que a menor diluição se encontra no quadrante 1 e a maior se encontra no quadrante 4. As diluições intermediárias seguem a ordem numérica, 0,05 no quadrante 2 e 0,005 no quadrante 3.

Para a curva de crescimento feita em microplaca na presença das diferentes concentrações de etanol, foi possível observar que em 0 % de etanol (v/v) as duas linhagens cresceram de forma similar na fase exponencial e atingiram a fase estacionária no mesmo tempo de incubação. Na presença de 12 % de etanol (v/v), a linhagem Y02999 se destacou conseguindo crescer até D.O.<sub>600</sub> próxima de 1,2, enquanto a BY4741 atingiu D.O.<sub>600</sub> próxima de 0,4, sendo que ambas começaram o crescimento com a D.O.<sub>600</sub> um pouco inferior a 0,2 (dados mostrados na figura 7). Para as concentrações de 14 %, 16 % e 18 % (v/v) ambas as linhagens não cresceram durante todo o período de incubação, iniciando o crescimento com D.O.<sub>600</sub> um pouco inferior a 0,2 e permanecendo com esta D.O.<sub>600</sub> ao final de 60 horas (dados não mostrados na figura 7).

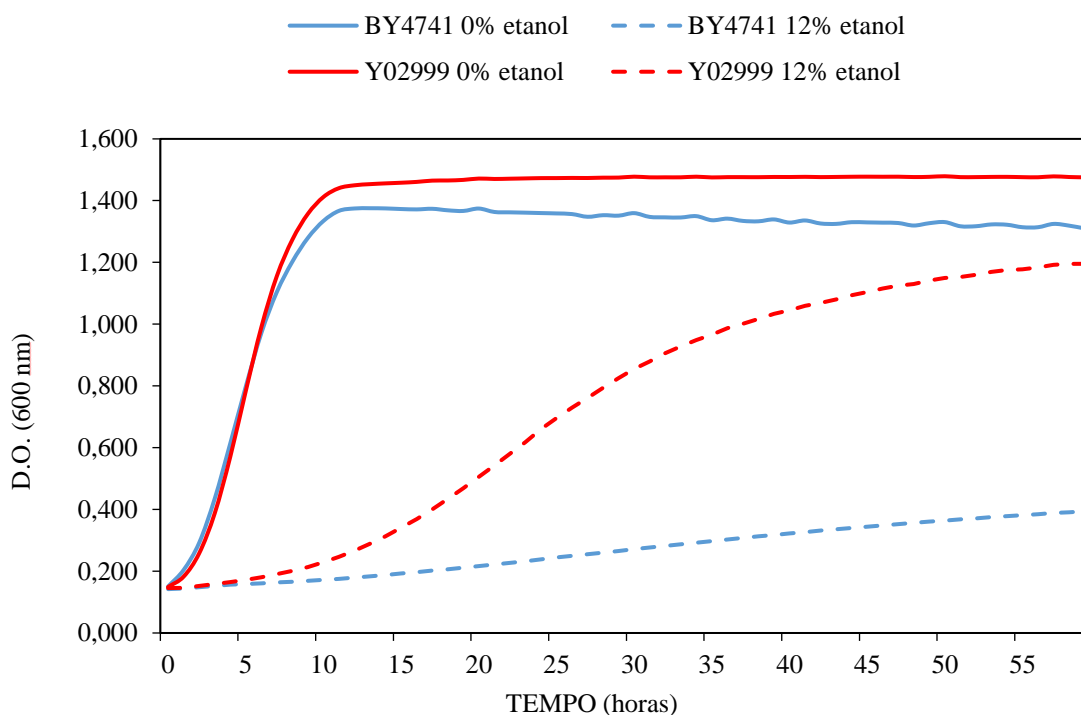


Figura 7: Curva de crescimento em microplaca na presença de etanol. Crescimento de *S. cerevisiae* BY4741 e Y02999 em meio YPD com diferentes concentrações de etanol [0 %, 12 %, 14 %, 16 % e 18 % (v/v)] após 60 horas de incubação a 30 °C. O ensaio foi realizado em sete replicatas técnicas para cada concentração de etanol e foi feita a média dos valores de D.O.<sub>600</sub> para plotagem do gráfico.

## 5.2 RNAseq

### 5.2.1 Análise da qualidade dos RNAs extraídos

A primeira etapa do experimento de RNAseq é a extração de RNA e este deve ser de boa qualidade para garantir o sucesso do sequenciamento. A figura 8 mostra a qualidade do RNA extraído para as linhagens de *S. cerevisiae* BY4741 e Y02999 na presença de 8% de etanol (v/v), tanto em gel de agarose quanto na análise pelo Bioanalyzer.

No gel é possível ver a integridade das bandas dos RNAs das subunidades 28 S e 18 S do ribossomo (rRNAs), além do 5 S e pequenos RNAs (sRNAs) na parte inferior do gel. O DNA contaminante que aparece no gel foi removido através do tratamento com DNase I. No Bioanalyzer é possível inferir a qualidade através da intensidade dos picos correspondentes aos rRNAs 28 S e 18 S, quanto maior o pico melhor a qualidade. Um acúmulo de picos entre 25 e 500 nucleotídeos [nt] pode significar degradação de rRNAs. Na figura 8 é possível ver que somente para os RNAs de *S. cerevisiae* BY4741 replicata 1 (BY4741\_1) houve um acúmulo mais pronunciado, diminuindo a intensidade dos picos referentes aos rRNAs 28 S e 18 S, porém este acúmulo não comprometeu a utilização desta replicata.

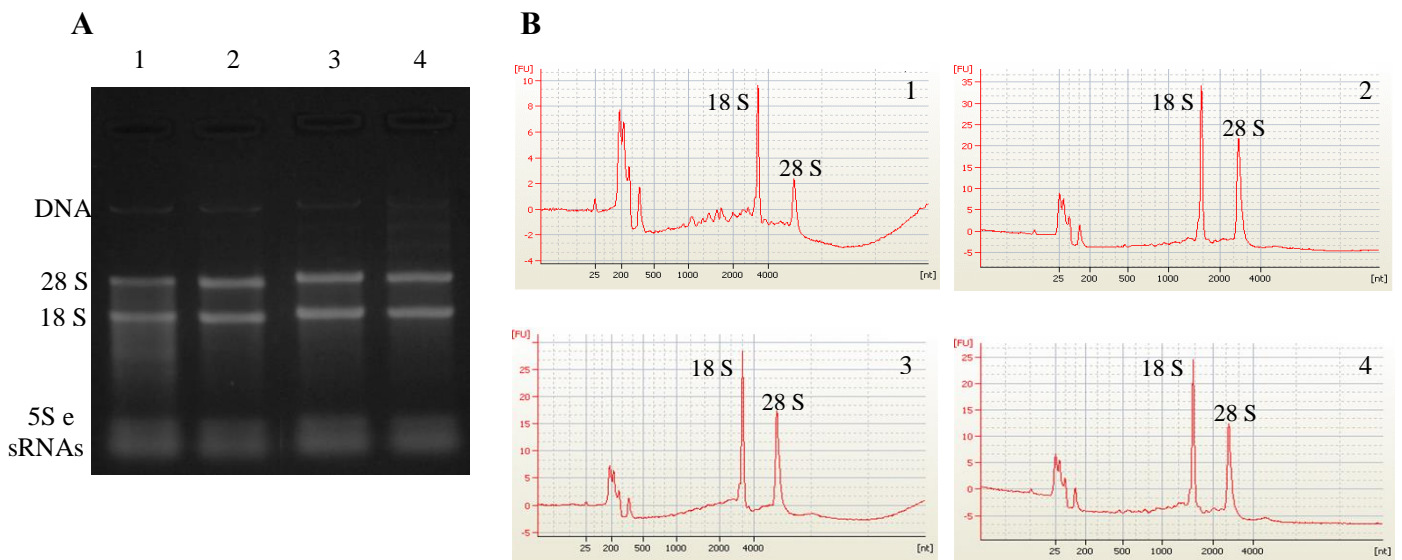


Figura 8: Análise da qualidade dos RNAs. **A.** Gel de agarose: 1- RNAs *S. cerevisiae* BY4741\_1; 2- RNAs *S. cerevisiae* BY4741\_2; 3- RNAs *S. cerevisiae* Y02999\_1; 4- RNAs *S. cerevisiae* Y02999\_2. **B.** Bioanalyzer: 1- RNAs *S. cerevisiae* BY4741\_1; 2- RNAs *S. cerevisiae* BY4741\_2; 3- RNAs *S. cerevisiae* Y02999\_1; 4- RNAs *S. cerevisiae* Y02999\_2.

### 5.2.2 Análise das bibliotecas construídas

A quantificação das bibliotecas em qPCR foi satisfatória para a quantidade de amostras necessárias para a injeção no sequenciador. A duplicata de *S. cerevisiae* BY4741 apresentou as concentrações de 59,41 nM e 140,93 nM, respectivamente e a duplicata de *S. cerevisiae* Y02999 apresentou 255,49 nM e 341,56 nM, respectivamente. O tamanho médio dos fragmentos obtidos foi calculado pelo Bioanalyzer. A figura 9 corresponde a um gráfico de distribuição destes fragmentos de acordo com o tamanho em pares de bases (pb). De acordo com a análise, o tamanho médio para as bibliotecas de BY4741 foram 368 pb e 387 pb, respectivamente, e para as bibliotecas de Y02999 foram 332 pb e 330 pb, respectivamente.

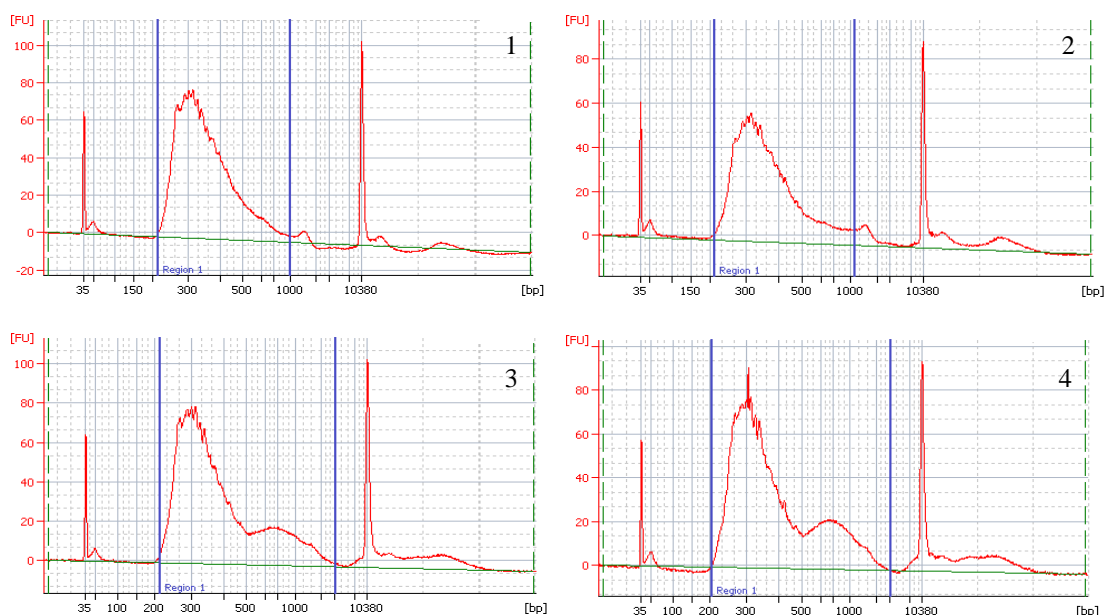
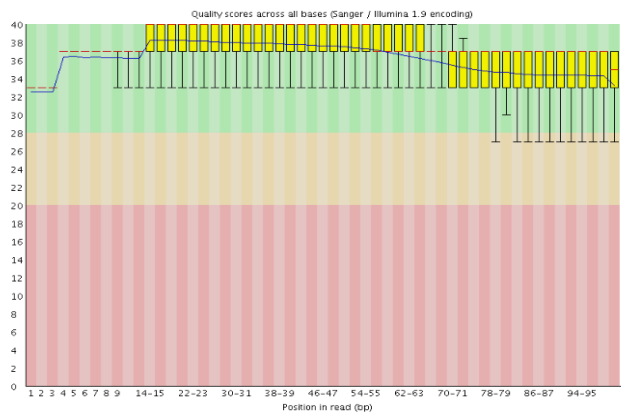
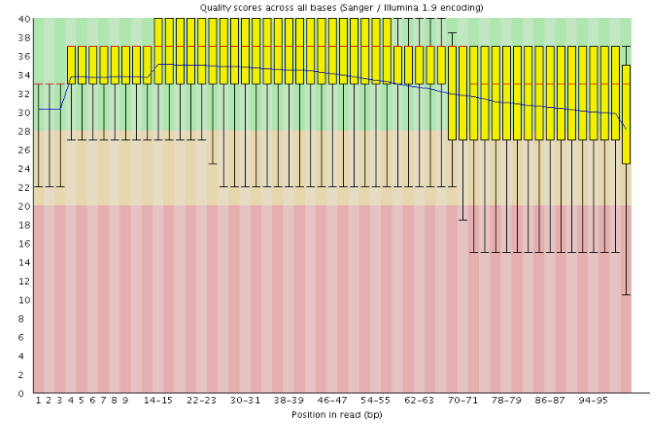
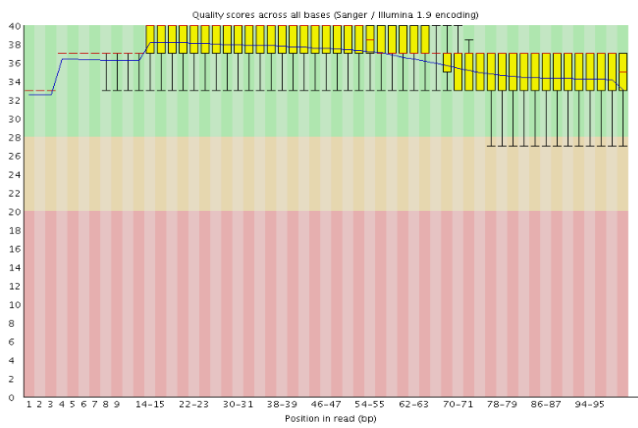
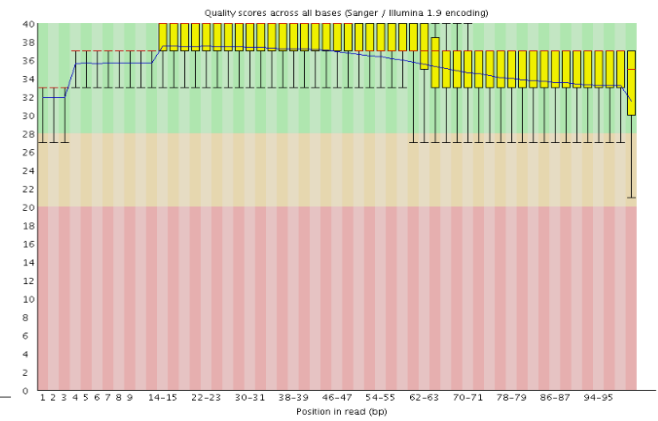
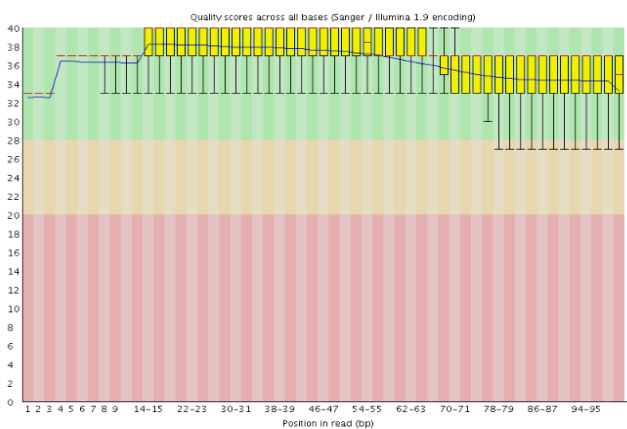
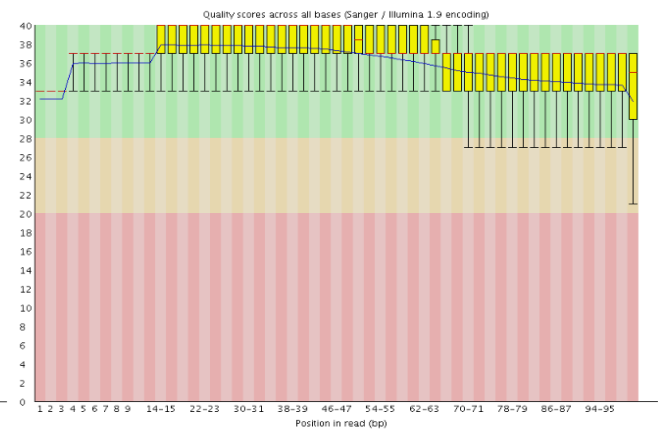


Figura 9: Análise do tamanho médio das bibliotecas em pb. 1- cDNAs ds *S. cerevisiae* BY4741\_1; 2- cDNAs ds *S. cerevisiae* BY4741\_2; 3- cDNAs ds *S. cerevisiae* Y02999\_1; 4- cDNAs ds *S. cerevisiae* Y02999\_2.

### 5.2.3 Análises de bioinformática

A análise de qualidade pelo programa FastQC indicou que a maior parte das *reads paired-end (forward e reverse)* geradas no sequenciamento estavam de alta qualidade, como mostrado na figura 10.

**A****BY4741\_1\_R1****BY4741\_1\_R2****B****BY4741\_2\_R1****BY4741\_2\_R2****C****Y02999\_1\_R1****Y02999\_1\_R2**

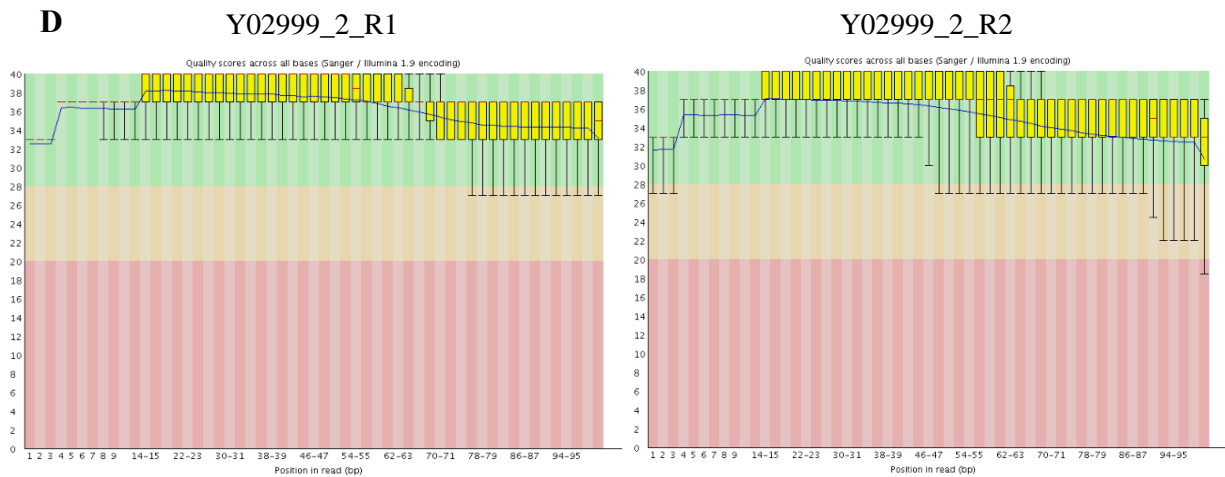
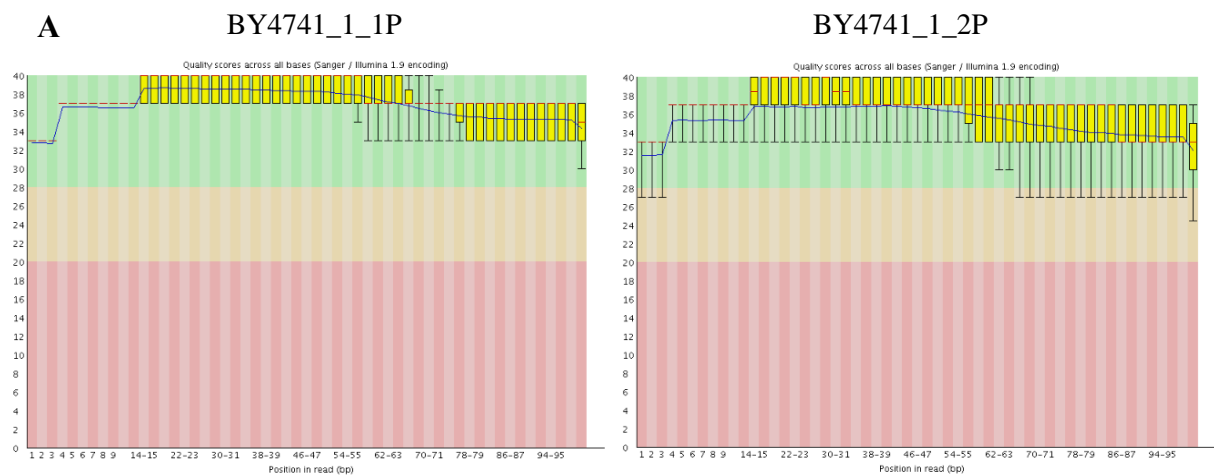


Figura 10: Pontuação de qualidade dos dados brutos obtidos no sequenciamento feita pelo FastQC. Como interpretação, temos que as *reads* contidas na faixa verde estão de alta qualidade, a faixa laranja representa as *reads* de qualidade intermediária e a faixa vermelha as *reads* de baixa qualidade. **A.** *Reads forward* (R1) e *reverse* (R2) da biblioteca BY4741\_1; **B.** Biblioteca BY4741\_2; **C.** Biblioteca Y02999\_1; **D.** Biblioteca Y02999\_2.

O Trimmomatic eliminou as sequências de baixa qualidade e como resultado do programa foram gerados quatro arquivos correspondentes às *reads forward* e *reverse* pareadas e não pareadas. Somente as *reads* pareadas foram utilizadas para fazer o alinhamento ao genoma de referência no programa TopHat e por isso foi feita uma nova análise de qualidade no FastQC somente com estas *reads*. A figura 11 ilustra o resultado obtido na segunda análise de qualidade.



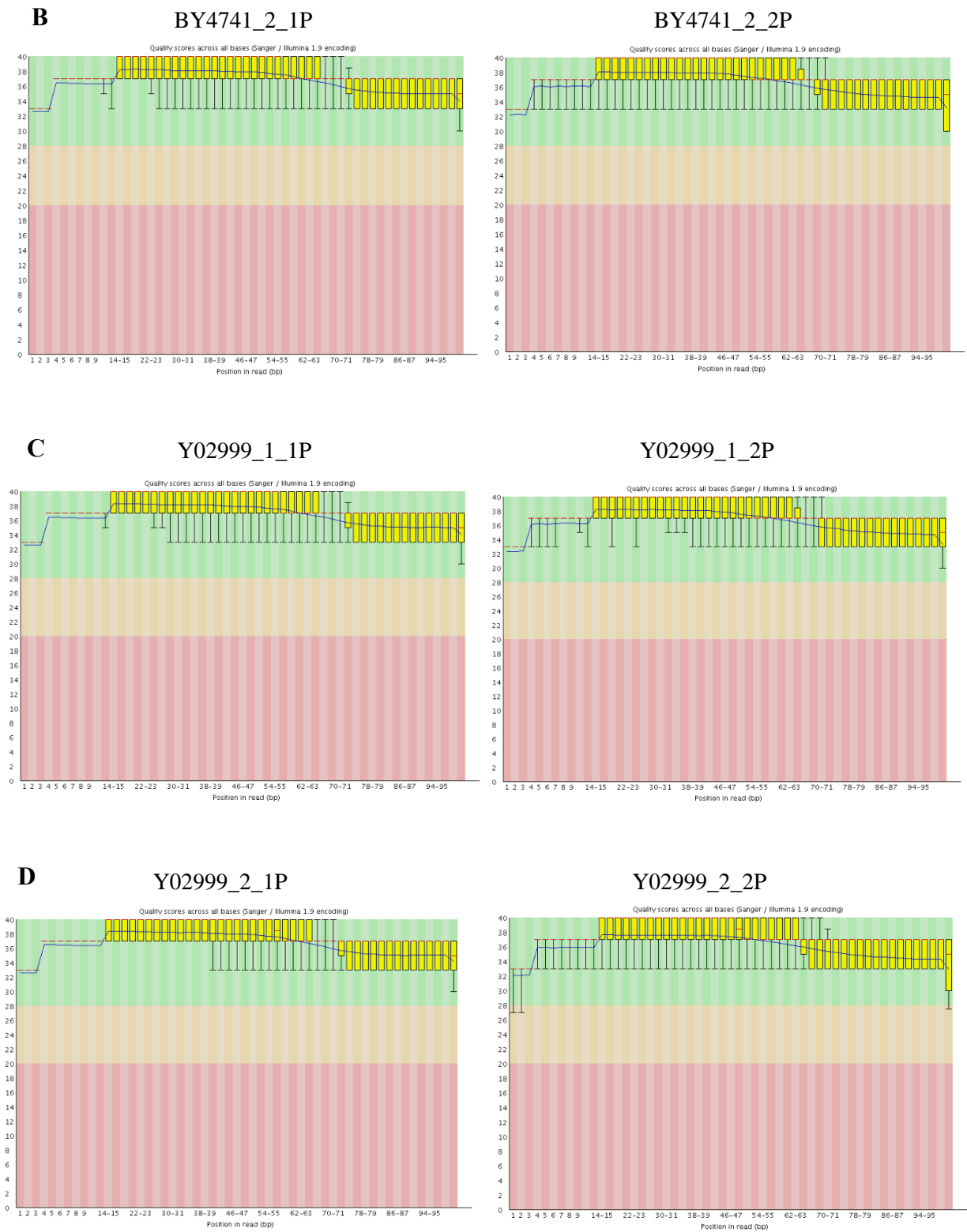


Figura 11: Pontuação de qualidade feita pelo FastQC das *reads* pareadas obtidas no Trimmomatic. **A.** *Reads forward* pareada (1P) e *reverse* pareada (2P) da biblioteca BY4741\_1; **B.** Biblioteca BY4741\_2; **C.** Biblioteca Y02999\_1; **D.** Biblioteca Y02999\_2.

A tabela 4 mostra algumas características das *reads* geradas no sequenciamento e indica o número de *reads* pareadas restantes após o corte de qualidade.

Tabela 4: Características das *reads* geradas no sequenciamento.

<i>S. cerevisiae</i>	Nº total de <i>reads</i> sequenciadas (milhões)	Tamanho médio da <i>read</i> (pb)	% GC	Nº total de <i>reads</i> pareadas após corte de qualidade (milhões)
BY4741_1	10.807.267	100	44	9.801.152
BY4741_2	10.614.994	100	44	10.337.367
Y02999_1	5.191.535	100	44	5.101.666
Y02999_2	7.824.432	100	44	7.551.733

A cobertura (C) do sequenciamento é um importante parâmetro a ser calculado pois ela se refere à média do número de *reads* que alinharam ao genoma de referência. Quanto maior este valor, significa que mais *reads* cobriram uma mesma base e o resultado do sequenciamento se torna mais confiável. O cálculo é feito utilizando o tamanho do genoma do organismo sequenciado (G) - que neste caso é 12 Mb, o tamanho médio das *reads* geradas no sequenciamento (T) e o número de *reads* que alinharam no genoma de referência (N), onde estas variáveis se relacionam de acordo com a equação  $C = N \times T / G$ . Neste trabalho, o número de *reads* considerado corresponde ao número de *reads* pareadas que restaram após o trimagem de qualidade (valores descritos na última coluna da tabela 4). Pela análise do alinhamento feito pelo TopHat foi observado que a porcentagem de *reads* pareadas que foram mapeadas foi maior que 95 % e por isso o número real de *reads* que alinharam é muito próximo ao número de *reads* pareadas que foram utilizadas pelo programa. Outro fator que deve ser considerado é o tipo *paired-end* do sequenciamento, implicando na multiplicação do tamanho da *read* por 2, já que cada uma delas foram lidas 100 pb no sentido *forward* e 100 pb no sentido *reverse*. A fórmula utilizada então foi:  $C = N \times T \times 2 / G$ .

Tabela 5: Cobertura aproximada calculada para as amostras sequenciadas

<i>S. cerevisiae</i>	Cobertura aproximada
BY4741_1	163 x
BY4741_2	173 x
Y02999_1	85 x
Y02999_2	126 x

A contagem das *reads* mapeadas no genoma de referência foi feita pelo Cufflinks. Ao analisar o resultado do alinhamento, foi visto que uma amostra da duplicata Y02999

apresentava contagem de *reads* para o gene *APJI*. Para verificar se as *reads* mapeadas para o gene *APJI* nesta amostra estava distribuída em toda a extensão do gene foi feita uma análise utilizando o programa Tablet (Milne *et al.*, 2010) e pelo resultado do programa conseguimos visualizar *reads* que cobriam todo o gene. Como a expressão de *APJI* em Y02999 foi bloqueada pela inserção de um gene de resistência a geneticina dentro da região codificadora de *APJI*, conseguimos concluir pelo resultado do programa que o gene *APJI* estava presente nesta amostra e por isso está replicata teve que ser eliminada da análise. A partir deste acontecimento, o cálculo de FC foi feito dividindo os valores de FPKM de cada gene da unicata Y02999 pela média dos valores de FPKM da duplicata de BY4741.

Como resultado, foram encontrados muitos genes com diferentes valores de  $\log_2$  FC. A partir deles foi feito um filtro para selecionar aqueles que seriam apresentados e o valor escolhido foi  $\log_2$  FC maior ou igual que 1,5 e menor ou igual que -1,5, ou seja, genes que estão pelo menos três vezes mais ou menos presentes em Y02999 em relação a BY4741 na condição de estresse por etanol submetida. Em números, foram encontrados 61 genes com valores de  $\log_2$  FC  $\geq$  que 1,5 (considerados genes *up*-regulados) e 41 genes com valor de  $\log_2$  FC  $\leq$  que -1,5 (genes *down*-regulados), totalizando 102 genes diferencialmente expressos entre as linhagens. Além destes, 18 genes apresentaram valores de FPKM somente para a *S. cerevisiae* Y02999 e 138 genes estavam presentes somente na linhagem *S. cerevisiae* BY4741. As tabelas com a descrição de cada um destes genes se encontram no **Anexo A**.

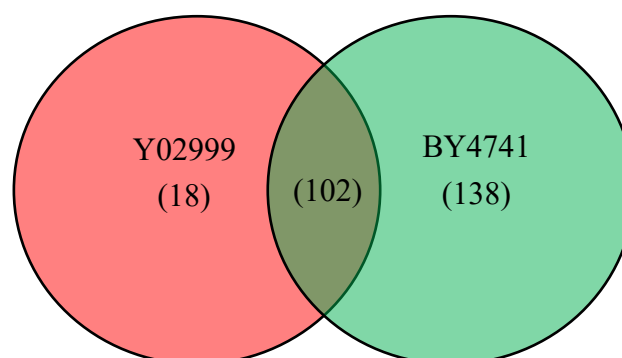


Figura 12: Diagrama de Venn ilustrando o resultado do sequenciamento. Foram encontrados 18 genes com valores de FPKM somente na linhagem Y02999, 138 genes ausentes em Y02999 e presentes em BY4741 e 102 genes diferencialmente expressos entre as linhagens, considerando  $\log_2$  FC  $\geq$  1,5 e  $\leq$  -1,5.

Devido a eliminação de uma amostra de Y02999 não foi possível comparar estatisticamente os valores de FPKM obtidos para as duas linhagens. Sem a análise estatística os resultados do sequenciamento não são tão confiáveis e torna-se indispensável a validação dos resultados por outra metodologia.

Quatro genes encontrados no RNAseq foram escolhidos para validação da expressão por qPCR. Destes, dois genes (*ARG3* e *BTN2*) que se encontravam *up*-regulados em Y02999 na presença do estresse de 8 % de etanol (v/v) e dois genes (*GMC2* e *COS12*) que não apresentaram FPKM em Y02999. Outros quatro genes (*OLE1*, *HSP104*, *UBC9* e *SIZ1*) foram incluídos na análise pelo qPCR. *OLE1* e *HSP104* foram escolhidos por serem genes descritos na literatura que exercem papéis importantes em linhagens de *S. cerevisiae* mais tolerantes ao etanol. *UBC9* e *SIZ1* foram escolhidos por serem componentes da maquinaria de sumoilação das células. Como *APJI* está relacionada a degradação de proteínas sumoiladas, surgiu o interesse em avaliar se existia um controle em nível transcricional da maquinaria de sumoilação na presença do etanol nas duas linhagens (BY4741 e Y02999). Apesar da escolha destes 4 últimos genes descritos serem baseadas somente na literatura, buscamos o valor de log<sub>2</sub> FC que eles apresentaram na análise pelo RNAseq. Estes valores estão mostrados na tabela 6.

Tabela 6: Valores de *Fold Change* (FC) e log<sub>2</sub> FC encontrados para os genes escolhidos para a análise no qPCR.

<b>Gene</b>	<b>FC</b>	<b>log<sub>2</sub> FC</b>
<i>ARG3</i>	3,0324	1,6005
<i>BTN2</i>	2,9377	1,5547
<i>OLE1</i>	1,4960	0,5811
<i>HSP104</i>	1,4130	0,4988
<i>UBC9</i>	1,3547	0,4380
<i>SIZ1</i>	1,2441	0,3151

#### 5.2.4 Análise dos genes escolhidos por qPCR

O experimento de qPCR foi feito com RNAs extraídos em triplicata biológica das duas linhagens em três concentrações de etanol, 0 %, 8 % e 12 % (v/v). Os géis apresentados abaixo representam a confirmação da síntese de todos os cDNAs utilizados no experimento por PCR convencional e ainda mostram a deleção de *APJI* em *S.*

*cerevisiae* Y02999 e a presença em BY4741. A deleção do gene em Y02999 foi confirmada por PCR convencional antes do ensaio em qPCR devido ao problema encontrado para uma amostra de Y02999 utilizada no RNAseq. O par de *primer* desenhado para *APJI* de *S. cerevisiae* anela em uma região dentro de *APJI* onde o gene de geneticina foi inserido durante a construção desta linhagem pela Euroscarf. Assim, não pode existir produto de PCR para estes *primers* quando a reação é realizada utilizando os cDNAs provenientes dos mRNAs extraídos dos cultivos de Y02999 como *template*. A inexistência deste produto de PCR significa que as amostras de Y02999 utilizadas no experimento são realmente silenciadas para *APJI*, como pode ser visto na figura 13.

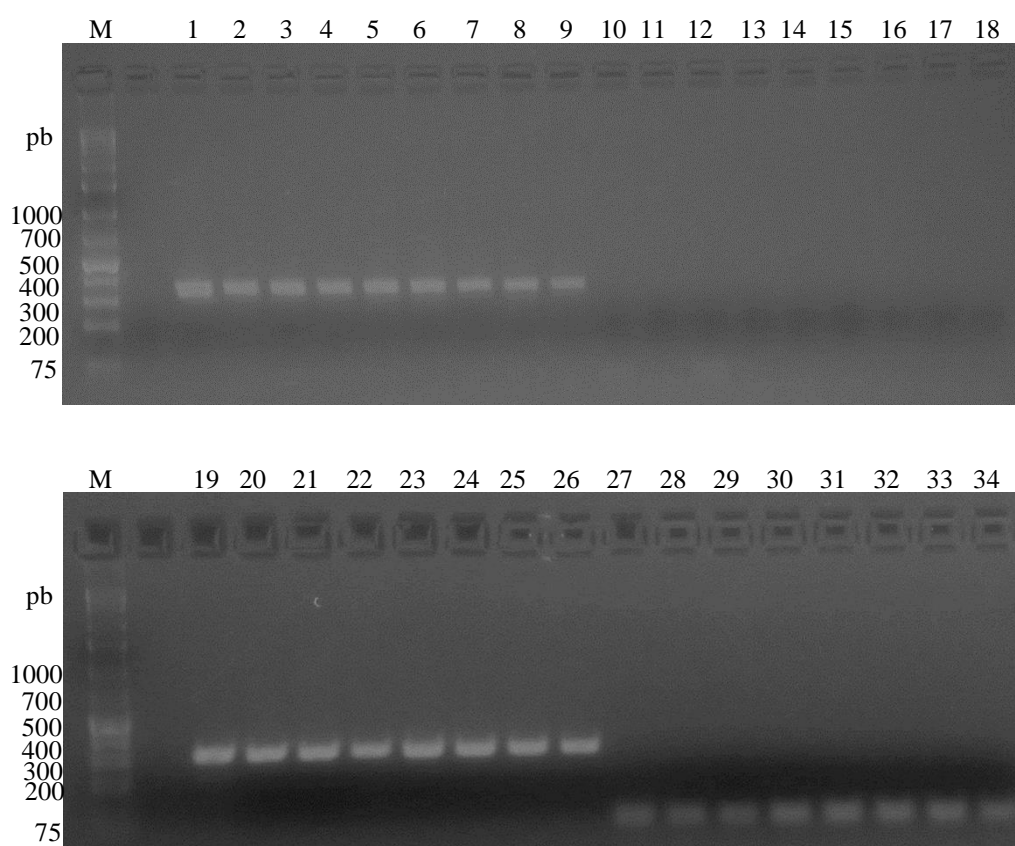
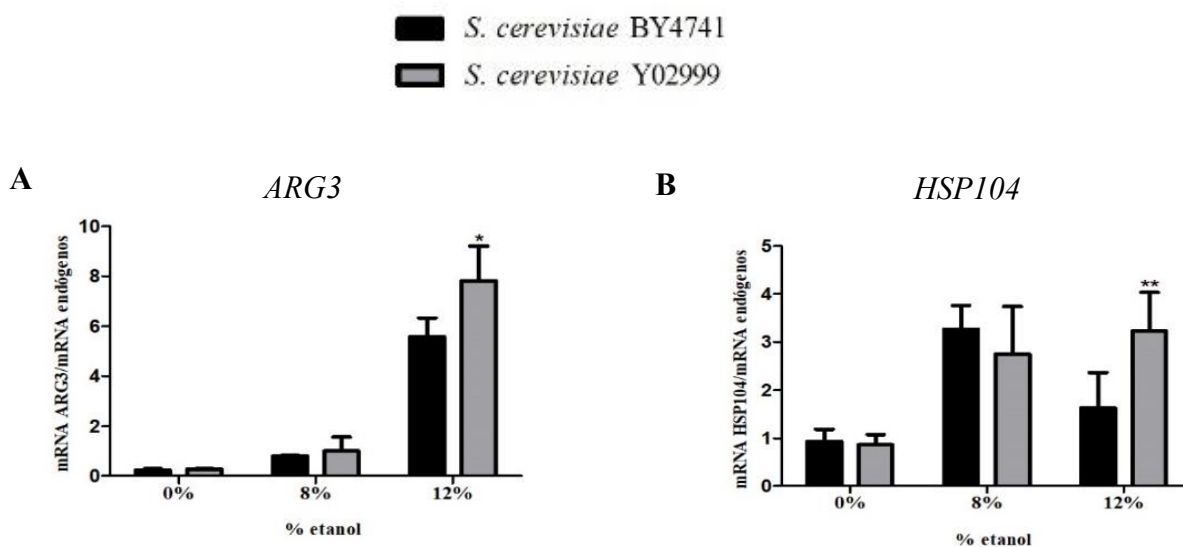


Figura 13: Gel de agarose para visualização da síntese de cDNA e confirmação da presença de *APJI* nos RNAs extraídos de *S. cerevisiae* BY4741 e ausência de *APJI* nos RNAs de *S. cerevisiae* Y02999 por PCR, ambas cultivadas na presença de 0 %, 8 % e 12 % de etanol (v/v). M: 1 Kb DNA Ladder (CELLCO); 1-9: *Primer IPP1* confirmando a síntese de cDNA para *S. cerevisiae* Y02999 nas três concentrações de etanol e em triplicata biológica; 11-18: *Primer APJI* confirmando a ausência deste gene em todos os cDNAs de *S. cerevisiae* Y02999; 19-26: *Primer IPP1* confirmando a síntese de cDNA para *S. cerevisiae* BY4741 nas três concentrações de etanol e em triplicata biológica para as concentrações 0 % e 8 % etanol (v/v) e duplicata biológica para

a concentração 12 % etanol (v/v). 27-34: *Primer APJ1* confirmando a presença deste gene em todos os cDNAs de *S. cerevisiae* BY4741;

Em 8 % de etanol (v/v) (concentração utilizada no RNAseq), apesar de todos os genes avaliados não terem apresentado diferenças de expressão significativas entre as linhagens (BY4741 e Y02999), é possível ver uma tendência de *up*-regulação em Y02999 para o gene *BTN2* (figura 14-D), o que vai de acordo com o resultado encontrado para este gene no RNAseq. Para os genes *COS12* (figura 14-G) e *GMC2* (figura 14-H) é possível observar uma tendência de *down*-regulação em Y02999. Este resultado também pode ser relacionado com o resultado obtido no RNAseq, onde estes genes se encontravam ausentes nesta linhagem. Na presença do estresse de 12 % de etanol (v/v), foi possível observar diferenças de expressão significativas entre as linhagens para os genes *ARG3*, *OLE1* e *HSP104*, onde *ARG3* (figura 14-A) e *HSP104* (figura 14-B) estão *up*-regulados em Y02999 e *OLE1* (figura 14-C) se mostrou *down*-regulado nesta linhagem. O gene *BTN2* (figura 14-D) apesar de não ter apresentado diferença significativa de expressão entre as linhagens em 12 % de etanol (v/v), manteve a tendência de *up*-regulação em Y02999 como observado em 8 % de etanol (v/v). O gene *COS12* (figura 14-G) manteve a tendência de *down*-regulação em Y02999 na presença 12 % de etanol (v/v). Já o gene *GMC2* (figura 14-H), apresentou uma tendência de *up*-regulação em Y02999 na presença de 12 % de etanol (v/v), diferentemente do observado em 8 % de etanol (v/v). Os genes *UBC9* (figura 14-E) e *SIZ1*(figura 14-F) não apresentaram nenhuma alteração de expressão entre as linhagens na presença das diferentes concentrações de etanol.



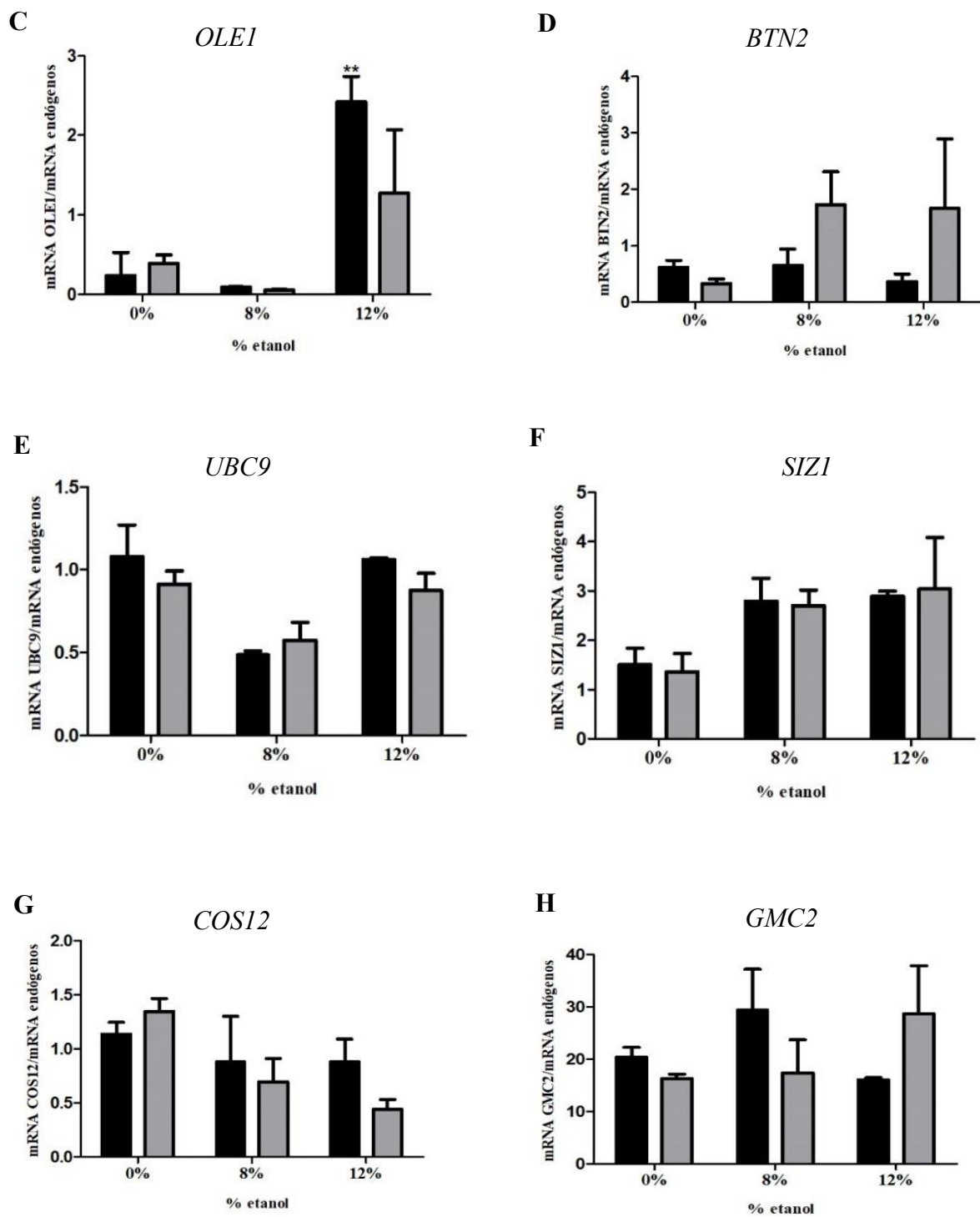


Figura 14: Análise de expressão em qPCR dos genes *ARG3*, *BTN2*, *COS12*, *GMC2*, *OLE1*, *HSP104*, *UBC9* e *SIZ1* para as linhagens *S. cerevisiae* BY4741 e Y02999 cultivadas nas três concentrações de etanol: 0 %, 8 % e 12% (v/v). **A.** *ARG3*; **B.** *HSP104*; **C.** *OLE1*; **D.** *BTN2*; **E.** *UBC9*; **F.** *SIZ1*; **G.** *COS12*; **H.** *GMC2*. O asterisco (\*) se refere aos valores das médias que são diferentes estatisticamente pelo teste de Bonferroni. (\*) p-valor < 0,05 e (\*\*) p-valor < 0,01.

## 5.3 CRISPR-Cas9

### 5.3.1 Clonagem do gRNA no pML104

Das colônias enviadas para sequenciamento, somente uma apresentava o gRNA, mas ele não foi inserido corretamente no vetor. A enzima *Ksp22 I* não digeriu o vetor, somente a *Swa I*. Sendo assim, o gRNA foi clonado somente no sítio de digestão da *Swa I*, entrando em duas cópias, como mostrado no resultado do sequenciamento de Sanger apresentada na figura 15.

```
5' CTTGGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGACAATCTTTGA
AAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGAACTTGATGTTTTCTTTTCGAGT
ATATACAAGGTGATTACATGTACGTTTGAAGTACAACCTCTAGATTTTGTAGTGCCCTCTGGGC
TAGCGGTAAAGGTGCGCATTTTTTCACACCCTACAATGTTCTGTTCAAAGATTTTGGTCAAAC
GCTGTAGAAGTGAAAGTTGGTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAA
ATGATCAGAATCGATGCATTTCTAGCTCTAAAACGAATATCGTAAAGTTTCGTAGATCTACGAA
ACTTTACGATATTCGTTTTAAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT
TGAAAAGTGGCACCGAGTCGGTGGTGCTTTTTGTTTTTAGTCAGCG 3'
```

Legenda:

TGATCA: Sítio de clivagem da *Ksp22 I*

ATTTAAAT: Sítio de clivagem *Swa I*

GATCTACGAACTTTACGATATTCGTTTTAAGCTAG: Oligo 1 (cópia1)

CTAGCTCTAAAACGAATATCGTAAAGTTTCGTA: Oligo 2 (cópia 2)

Figura 15: Clonagem de duas cópias do gRNA no sítio de clivagem da *Swa I* no vetor pML104.

Ao investigar o motivo do não funcionamento da *Ksp22 I*, descobriu-se que esta enzima não digere DNAs metilados, e como a *E. coli* DH5 $\alpha$  apresenta este sistema de metilação, foi necessário a busca por uma outra linhagem de bactéria que apresentasse este sistema bloqueado para replicar o vetor e em seguida conduzir a digestão. Conseguiu-se a linhagem de *E. coli* GM2163 e novas tentativas de digestão e clonagem foram realizadas como descrito no item 4.4.1, mas todas sem sucesso. Apesar das duas enzimas utilizadas na digestão estarem aparentemente funcionais (figura 16), todas as colônias transformantes investigadas se apresentavam na PCR com os *primers* T3R e seqF da mesma forma que a PCR do vetor vazio (sem gRNA), como mostrado na figura 17.

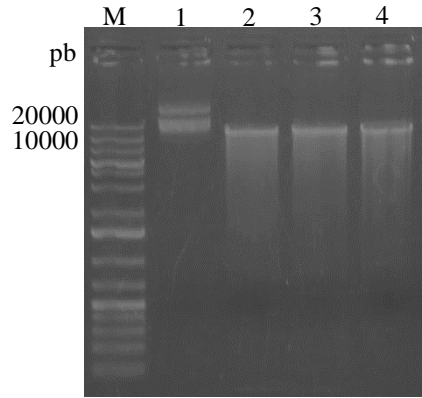


Figura 16: Gel de agarose 0,8 % mostrando o resultado da digestão de pML104 extraído da *E. coli* GM2163. M: 1 kb DNA Ladder Plus (Fermentas); 1: pML104 não digerido; 2: pML104 digerido *Swa* I; 3: pML104 digerido *Ksp22* I; 4: pML104 digerido simultaneamente com as duas enzimas.

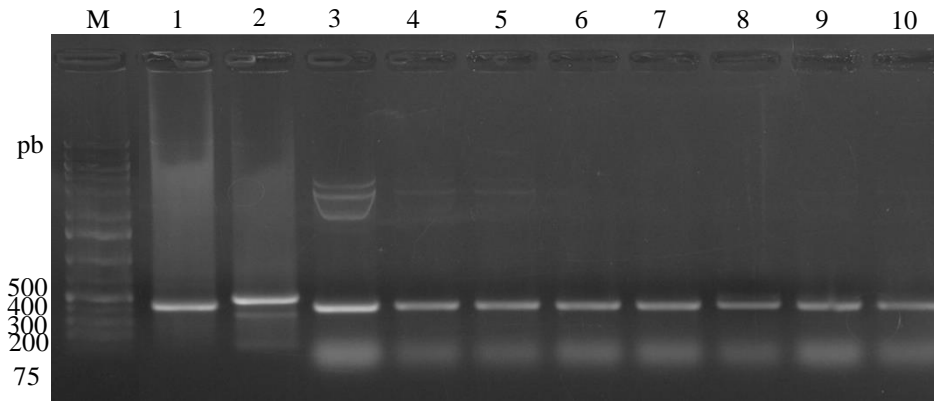


Figura 17: Gel de agarose 1 % ilustrando o resultado da PCR com *primers* T3R e seqF de 8 colônias transformantes obtidas da transformação de DH5α com a reação de ligação entre o híbrido (oligo 1 e oligo 2) e o vetor pML104 (não metilado) linearizado após a digestão com as enzimas *Ksp22* I e *Swa* I. M: 1 kb DNA Ladder (CELLCO); 1: PCR pML104 vazio; 2: PCR pML104 2 cópias de gRNA; 3-10: PCR das 8 colônias investigadas para a presença do gRNA.

Com isso, decidiu-se transformar a *K. marxianus* CCT7735  $\Delta$ *ura3* com a construção pML104-2guias. Apesar do vetor estar com duas cópias do gRNA, o sentido da clonagem no vetor está correto e possivelmente este gRNA pode guiar a edição do alvo.

### 5.3.2 Validação da expressão de Cas9

Após a extração de RNA das leveduras *S. cerevisiae* AH109 e *K. marxianus* CCT735  $\Delta$ *ura3* transformadas com o vetor pML104 e não transformadas, foi feito o tratamento dos RNAs com DNase I e a eficácia do tratamento foi confirmada por PCR. O resultado do tratamento está mostrado no gel de agarose correspondente à figura 18.

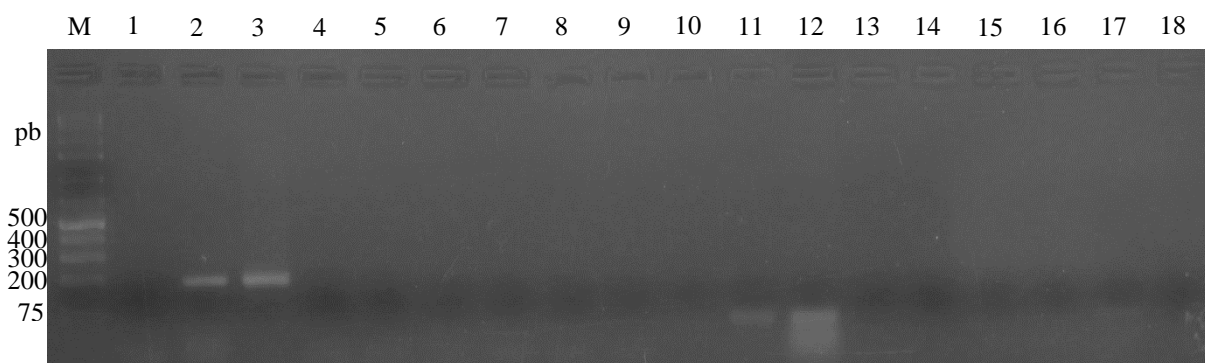


Figura 18: Gel de agarose 2% mostrando o resultado do tratamento dos RNAs de *S. cerevisiae* AH109 e *K. marxianus* CCT7735  $\Delta$ *ura3* com DNase I. M: 1 kb Ladder (CELLCO); 1: Controle negativo (C-) *K. marxianus* - primers gene *ACT1* e água ao invés de *template*; 2-3: Controle positivo (C+), DNA genômico *K. marxianus* como *template*; 4-9: RNAs *K. marxianus*; 10: C- *S. cerevisiae* - primers gene *ACT1* e água ao invés de *template*, 11-12: C+, DNA genômico *S. cerevisiae*, 13-18: RNAs *S. cerevisiae*.

Em seguida, foi feita a síntese de cDNA para todas as linhagens e condições do experimento. A confirmação da síntese foi feita por PCR e o resultado está mostrado no gel abaixo (figura 19).

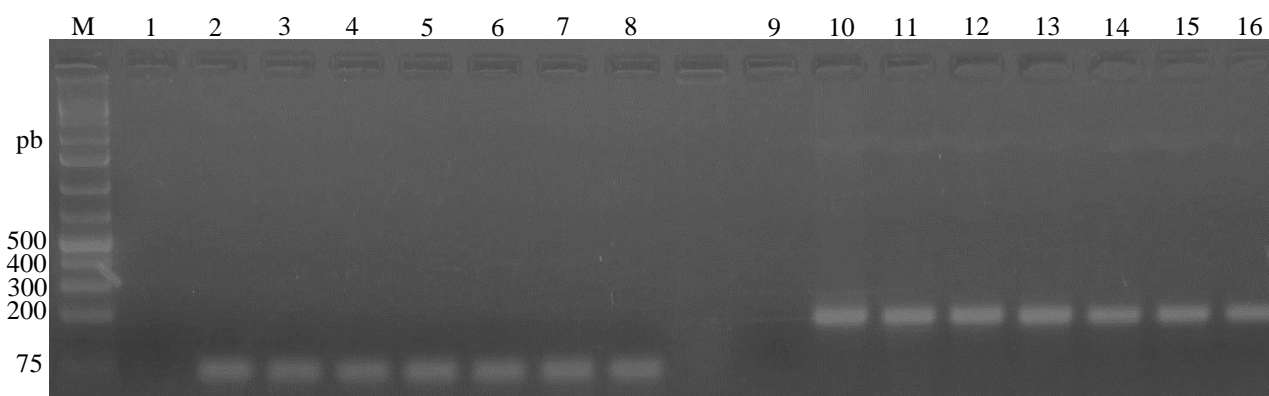


Figura 19: Gel de agarose 2% mostrando a síntese de cDNAs a partir dos RNAs extraídos de *S. cerevisiae* AH109 e *K. marxianus* CCT7735  $\Delta$ *ura3*. M: 1 kb Ladder (CELLCO); 1: C-; 2: C+; 3-8: PCR com primers gene *ACT1* *S. cerevisiae* usando os cDNAs como *template*; 9: C-; 10: C+; 11-16: PCR com primers gene *ACT1* *K. marxianus* usando os cDNAs como *template*.

Confirmado a qualidade dos cDNAs, foi então conduzido o experimento de validação da expressão de Cas9 em *S. cerevisiae*, microrganismo para qual o vetor pML104 foi desenvolvido, e confirmação da sua possível expressão em *K. marxianus*. Os controles do experimento foram as duas leveduras não transformadas com o vetor. Primeiramente foi desenhado o par de *primer* 1 (sequência no item 4.4.3) e estes foram testados no genoma das leveduras. Foi observado para *S. cerevisiae* que eles anelavam no genoma e gerava um *amplicon* do mesmo tamanho que o *amplicon* esperado para amplificação dentro da Cas9. Para *K. marxianus* este *primer* anelava somente no genoma da levedura transformada, o que era esperado devido a extração de vetor juntamente com o genoma da levedura (Figura 20-A).

Ao realizar o PCR com os cDNAs, observou-se bandas para todas as amostras, tanto das leveduras transformadas quanto não transformadas, além de uma contaminação no controle negativo (Figura 20-B).

O resultado obtido não é confiável devido a contaminação do controle negativo e mesmo fazendo novas diluições do estoque destes *primers*, o resultado permaneceu e não foi possível concluir sobre a expressão da Cas9 nas duas leveduras.

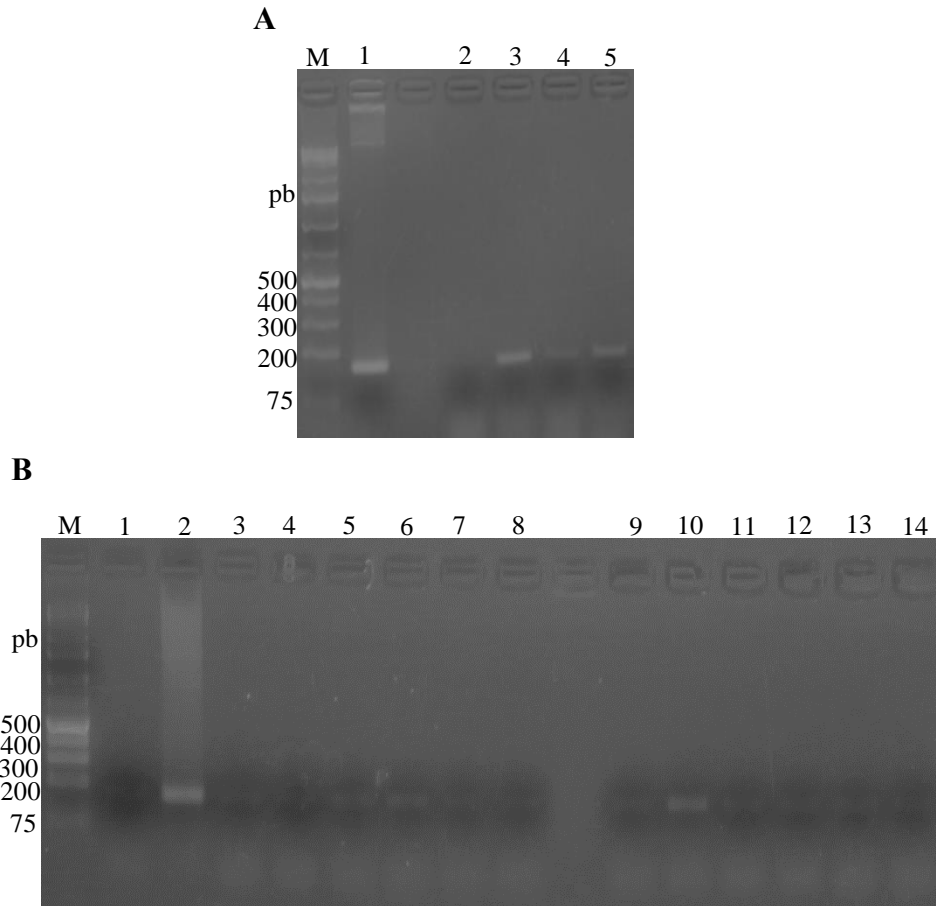


Figura 20: **A.** Gel de agarose 2 % mostrando o resultado do PCR com o par de *primer 1* para Cas9 usando o genoma das leveduras como *template*. M: 1 kb Ladder (CELLCO); 1: C+ (pML104); 2: *K. marxianus* não transformada; 3: *K. marxianus* transformada; 4: *S. cerevisiae* não transformada; 5: *S. cerevisiae* transformada. **B.** Gel de agarose 2% mostrando o resultado do PCR com o par de *primer 1* usando os cDNAs das duas leveduras como *template*. M: 1 kb Ladder (CELLCO); 1: C-; 2 C+; 3-5: *S. cerevisiae* não transformada; 6-8: *S. cerevisiae* transformada; 9-11: *K. marxianus* não transformada; 12-14: *K. marxianus* transformada.

Foi desenhado um novo par de *primer* (par2, sequência no item 4.4.3). Este foi testado nos cDNAs das duas leveduras e o mesmo resultado visto anteriormente se reproduziu, sendo que agora o controle negativo não apresentou amplificação (Figura 21-B). Ao testar estes *primers* no genoma das leveduras, novamente foi observado que eles anelavam no genoma das leveduras não transformadas e o *amplicon* era do mesmo tamanho esperado para a amplificação destes no gene da Cas9 contido no vetor (Figura 21-A).

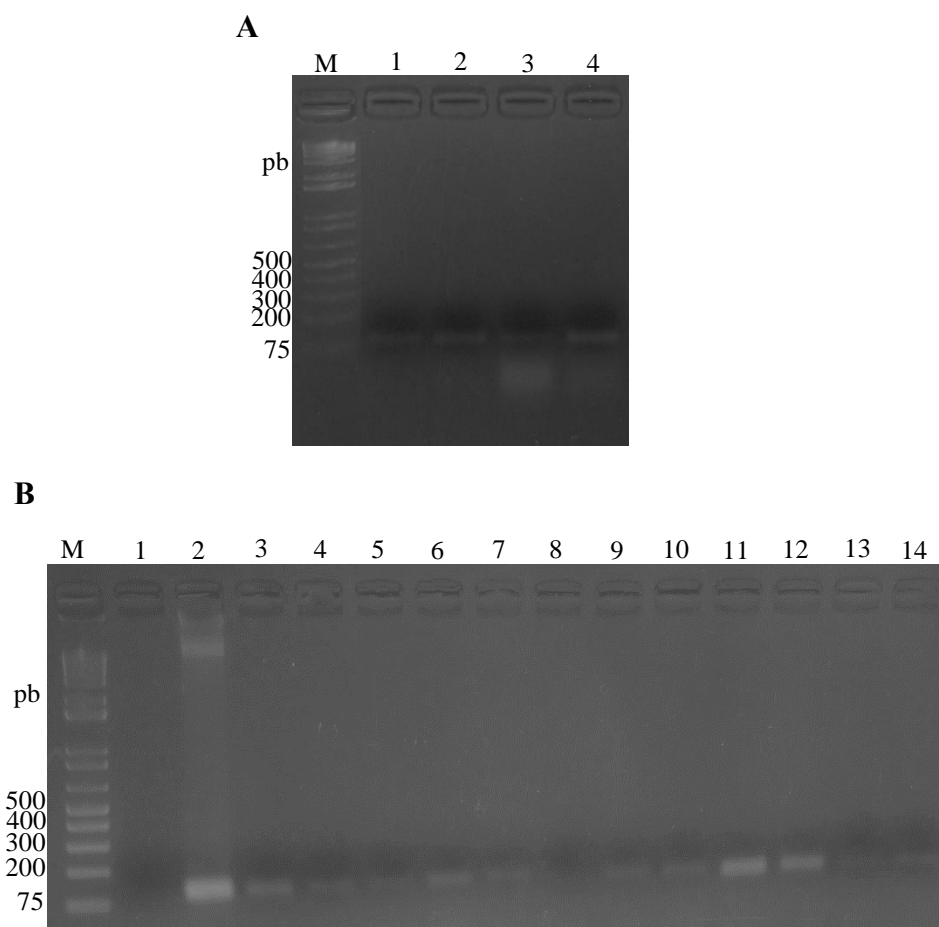


Figura 21: **A.** Gel de agarose 2 % mostrando o resultado do PCR com o par de *primer 2* para Cas9 usando o genoma das leveduras como *template*. M: 1 kb Ladder (CELLCO); 1: *K. marxianus* não transformada; 2: *K. marxianus* transformada; 3: *S. cerevisiae* não transformada; 4: *S. cerevisiae* transformada. **B.** Gel de agarose 2% mostrando o resultado do PCR com o par de *primer 2* usando os cDNAs das duas leveduras como *template*. M: 1 kb Ladder (CELLCO); 1: C-; 2 C+ (pML104); 3-5: *K. marxianus* não transformada; 6-8: *K. marxianus* transformada; 9-11: *S. cerevisiae* não transformada; 12-14: *S. cerevisiae* transformada.

Nas análises de bioinformática, tanto o par 1 como o par 2 estavam desenhados corretamente e não anelavam em pontos tão próximos no genoma das leveduras em questão. Considerando que se trata de *amplicons* muito pequenos (120 e 160 pares de bases, respectivamente), as chances destes anelamentos inespecíficos acontecerem eram muito pequenas.

Os resultados do experimento não foram conclusivos, porém sabendo que os *primers* anelam corretamente no pML104 (como mostrado nos controles positivos das figuras 20 e 21) e sabendo da ocorrência das ampliações inespecíficas no genoma das

leveduras, esperava-se que o resultado da PCR com os *primers* da Cas9 resultasse em bandas mais acentuadas para os cDNAs das leveduras transformadas caso a Cas9 estivesse sendo expressa (já que a DNA polimerase teria mais molde para amplificar) e bandas mais fracas caso a Cas9 não estivesse sendo expressa.

### 5.3.3 Busca por transformantes positivos $\Delta apj1$

Após a transformação de *K. marxianus* CCT7735  $\Delta ura$  com pML104-2guias, as colônias escolhidas para realização do sequenciamento de parte de *APJ1* foram submetidas à extração de DNA plasmidial para confirmação da construção presente na levedura. Foi feito um PCR com os *primers* T3R e seqF e o resultado está mostrado na figura 22.

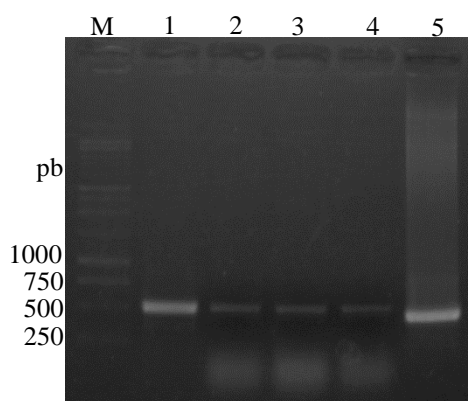


Figura 22: Gel de agarose 1% ilustrando o PCR com *primers* T3R e seqF das três colônias de *K. marxianus* transformadas com pML104-2 guias que tiveram parte de *APJ1* sequenciada. M: 1 kb Ladder Plus (Promega); 1: pML104-2guias; 2-4: vetores extraídos das 3 colônias transformantes; 5: pML104 vazio.

Foram enviadas 4 colônias para o sequenciamento Sanger, as três apresentadas no gel acima e mais um transformante não mostrado. O resultado do sequenciamento está apresentado na figura 23. Nenhuma das 4 colônias apresentaram alteração na sequência de nucleotídeos na região esperada, entre o ATG inicial e o TGG.

Colônia referência- *K. marxianus*  $\Delta$ ura

ATTCCCGAAGTATGGTGAAAGATACGAACTTTACGATATTCTGGAGGTGAGCGCGGAT  
GCCACTGATCAGGAGATCAAAAAGCGTACAGGAAAAGGCGTTGCGGCACCATCCAGA  
CAAAAACAACCATAGCCCAGAGTCTATAAAGTTATTCCAGGAATTAAGTCATGCGTATG  
AAATATTAAGTGATGAAAGTAAGAGAGAGTTGTACGATCAGTATGGGACTGTGGATGAG  
ACTGAAATCAATGAGATGATGTCTAAGCAAAGAGGTGCAGCTAGCAATGCGGCGCTTGC  
TCACACAGCAGGGGACCTTTTTGCGCAGTTTTTGGCGGCCGTCCGGGTCTGTTGTGCGA

Colônia 1: *K. marxianus*  $\Delta$ ura + pML104-2guias

ATTCCCGAAGTATGGTGAAAGATACGAACTTTACGATATTCTGGAGGTGAGCGCGGATG  
CCACTGATCAGGAGATCAAAAAGCGTACAGGAAAAGGCGTTGCGGCACCATCCAGAC  
AAAAACAACCATAGCCCAGAGTCTATAAAGTTATTCCAGGAATTAAGTCATGCGTATGA  
AATATTAAGTGATGAAAGTAAGAGAGAGTTGTACGATCAGTATGGGACTGTGGATGAGA  
CTGAAATCAATGAGATGATGTCTAAGCAAAGAGGTGCAGCTAGCAATGCGGCGCTTGTCT  
CACACAGCAGGGGACCTTTTTGCGCAGTTTTTGGCGGCCGTCCGGTCTGTTAG

Colônia 2: *K. marxianus*  $\Delta$ ura + pML104-2guias

ATCCAGAAGTATGGTGAAAGATACGAACTTTACGATATTCTGGAGGTGAGCGCGGATG  
CCACTGATCAGGAGATCAAAAAGCGTACAGGAAAAGGCGTTGCGGCACCATCCAGAC  
AAAAACAACCATAGCCCAGAGTCTATAAAGTTATTCCAGGAATTAAGTCATGCGTATGA  
AATATTAAGTGATGAAAGTAAGAGAGAGTTGTACGATCAGTATGGGACTGTGGATGAGA  
CTGAAATCAATGAGATGATGTCTAAGCAAAGAGGTGCAGCTAGCAATGCGGCGCTTGTCT  
CACACAGCAGGGGACCTTTTTGCGCAGTTTTTGGCGGCCGTCCGAGTCTGCTATGGGCG

Colônia 3: *K. marxianus*  $\Delta$ ura + pML104-2guias

ATGGGGGATCCAAAGTATGGTGAAAGATACGAACTTTACGATATTCTGGAGGTGAGCG  
CGGATGCCACTGATCAGGAGATCAAAAAGCGTACAGGAAAAGGCGTTGCGGCACCAT  
CCAGACAAAAACAACCATAGCCCAGAGTCTATAAAGTTATTCCAGGAATTAAGTCATGC  
GTATGAAATATTAAGTGATGAAAGTAAGAGAGAGTTGTACGATCAGTATGGGACTGTGG  
ATGAGACTGAAATCAATGAGATGATGTCTAAGCAAAGAGGTGCAGCTAGCAATGCGGCG  
CTTGTCTCACACAGCAGGGGACCTTTTTGCGCAGTTTTTGGCGGCCGTCCCGGGGTCTGC  
TTTG

Colônia 4: *K. marxianus*  $\Delta$ ura + pML104-2guias

ATGGGGGATACAGAAGTATGGTGAAAGATACGAACTTTACGATATTCTGGAGGTGAGC  
GCGGATGCCACTGATCAGGAGATCAAAAAGCGTACAGGAAAAGGCGTTGCGGCACCA  
TCCAGACAAAAACAACCATAGCCCAGAGTCTATAAAGTTATTCCAGGAATTAAGTCATG  
CGTATGAAATATTAAGTGATGAAAGTAAGAGAGAGTTGTACGATCAGTATGGGACTGTG  
GATGAGACTGAAATCAATGAGATGATGTCTAAGCAAAGAGGTGCAGCTAGCAATGCGGC  
GCTTGTCTCACACAGCAGGGGACCTTTTTGCGCAGTTTTTGGCGGCCGTCCGGGGTCTGC  
TAG

Figura 23: Resultado sequenciamento de Sanger de parte do gene *APJI* de *K. marxianus*. A alteração de nucleotídeos (inserções ou deleções) era esperada seria na região entre o ATG inicial e o TGG. Comparando as 4 colônias com a sequência referência, nenhuma alteração foi vista.

## 6. DISCUSSÃO

### 6.1 Influência da deleção de *APJ1* no transcriptoma de *S. cerevisiae* sob estresse por etanol

A hipótese deste trabalho começou a ser elaborada a partir de uma publicação de Swinnen *et al.* (2012), que mostrou o impacto da deleção de *APJ1* na tolerância a altas concentrações de etanol, na linhagem laboratorial *S. cerevisiae* BY4741 e em um segregante de uma linhagem industrial, chamada VR1- 5B. Em meio sólido contendo YP e etanol como única fonte de carbono, BY4741  $\Delta apj1$  cresceu melhor que a linhagem não deletada na presença de 14 % (v/v) de etanol. A VR1- 5B  $\Delta apj1$  apresentou melhor crescimento em 16 % (v/v) de etanol comparada à linhagem não deletada.

A definição de tolerância é muito variável e no trabalho de Swinnen *et al.* (2012) acredita-se que a tolerância à altas concentrações de etanol só é relevante no final da fermentação, quando o nível de açúcar do meio está muito baixo, por isso os experimentos foram conduzidos contendo etanol como única fonte de carbono.

Este experimento foi repetido neste trabalho, comparando o crescimento da linhagem BY4741 e a linhagem Y02999 ( $\Delta apj1$ ) em diferentes concentrações de etanol. Porém, o ensaio foi conduzido em meio YPD, tendo a glicose como principal fonte de carbono e adicionando as diferentes concentrações de etanol, assim como foi realizado nos trabalhos de Kasavi *et al.* (2016) e Li *et al.* (2017). O acréscimo de açúcar foi feito uma vez que, em uma dorna de fermentação, a concentração de açúcares é próxima de zero quando o processo está chegando ao final. O resultado obtido foi similar e Y02999 cresceu melhor que a BY4741 em 16 e 18 % de etanol (v/v) mesmo na presença da glicose. O artigo mostra o impacto da deleção somente em 14 % (v/v) e neste trabalho foi possível ver que a Y02999 consegue crescer melhor em concentrações ainda maiores.

Além de meio sólido, foi feita uma curva de crescimento em microplaca com meio YPD e diferentes concentrações de etanol e novamente foi possível validar o impacto da deleção pelo melhor crescimento de Y02999 em 12 % de etanol (v/v) em relação a BY4741. Neste experimento foi observado que quando se trata de crescimento em meio líquido o comportamento das linhagens são diferentes, uma vez que ambas não cresceram em 14 %, 16 % e 18 % (v/v). A explicação para esta diferença de crescimento entre meios, é que o meio líquido favorece o contato com o etanol aumentando o dano causado por este agente às leveduras.

O impacto da deleção de *APJI* em Y02999 foi comprovado em meio YPD sólido e líquido na presença de diferentes concentrações de etanol. Porém, pouco se sabe como a deleção deste gene promove o aumento da tolerância. Sendo assim, este trabalho propôs uma análise comparativa do transcriptoma da linhagem deletada e não deletada para este gene na presença do etanol, para entender se a deleção implica na alteração do perfil de transcrição da célula sem *APJI* e quem são esses genes diferencialmente expressos que contribuem para o aumento da tolerância. A análise do transcriptoma de BY4741 e Y02999 foi conduzida após as células sofrerem o estresse de 8 % de etanol (v/v) por 2 horas.

Como resultado do RNAseq foram encontrados 102 genes diferencialmente expressos entre as linhagens, considerando  $\log_2 FC \geq 1,5$  ou  $\leq -1,5$ , ou seja, genes que estavam três vezes mais ou menos presentes em Y02999 em relação à cepa parental BY4741 na condição de estresse com 8 % de etanol (v/v). Além destes, foram encontrados 18 genes presentes somente em Y02999 e 138 genes presentes somente em BY4741. No Anexo A deste trabalho está a tabela descrevendo todos estes genes.

- Genes *up*-regulados em Y02999

Entre os genes mais expressos, estão genes de funções desconhecidas, genes envolvidos na meiose, esporulação, biossíntese de parede celular, dentre outros processos biológicos. Três genes *up*-regulados chamaram a atenção por serem genes correlacionados a tolerância ao etanol, segundo informações da literatura: *VARI*, *ARG3* e *BTN2* (Litvinchuk *et al.*, 2013; Cheng *et al.*, 2013; Espinazo-Romeu *et al.*, 2008).

*VARI* é uma proteína presente na subunidade menor dos ribossomos mitocondriais que se apresentou 3,2749 vezes mais expresso em Y02999. Um trabalho de Litvinchuk *et al.* (2013) mostrou que essa proteína se liga ao DNA mitocondrial após uma condição de estresse e pode promover quebras no DNA ou interferir na replicação. Mesmo na ausência de DNA mitocondrial a levedura consegue crescer e um possível benefício da perda é a diminuição da produção de ROS durante a condição de estresse. ROS são compostos tóxicos que causam danos severos em diversos componentes celulares, resultando na oxidação de proteínas, peroxidação de lipídeos e oxidação do DNA tanto mitocondrial como cromossomal (Auesukaree, 2017). Outros genes responsivos a estresse oxidativo como *SOD1*, *SOD2*, *YAP1*, *SKN7*, *GSH1*, *GSH2*, *TRX1* e *TRR1* (Auesukaree, 2017), estavam presentes nas duas linhagens, mas não se apresentaram diferencialmente expressos. *AIM11* é um gene que quando deletado leva ao aumento da perda de DNA

mitocondrial em leveduras (Hess *et al.*, 2009). Neste trabalho, ele estava 3,654 vezes mais presente na linhagem menos tolerante, BY4741, ou seja, era um gene *down*-regulado em Y02999. Sabendo que tanto *VARI* como *AIM11* estão envolvidos na perda de DNA mitocondrial, hipotetiza-se que a menor expressão de *AIM11* em Y02999 esteja contribuindo para a perda de DNA mitocondrial e contribuindo com a atividade de *VARI*. Não foi possível validar a expressão de *VARI* por qPCR, pois como se trata de um gene mitocondrial sua sequência é rica em adenina (A) e timina (T) e o *primer* que foi desenhado para validação não funcionou na condição que o experimento foi conduzido. Novos *primers* estão sendo sintetizados para realização do experimento.

*ARG3* estava 3,0324 vezes mais expresso em Y02999 e corresponde a ornitina carbamoiltransferase, enzima que catalisa a biossíntese de citrulina a partir da ornitina, sendo a citrulina um precursor para biossíntese de arginina. Cheng *et al.* (2016) mostraram o efeito protetivo da arginina contra o estresse por etanol em linhagens de *S. cerevisiae* que superexpressavam *ARG4*, gene que codifica a arginina succinato liase, enzima que catalisa a última etapa da biossíntese de arginina. A superexpressão resultava em um grande acúmulo de arginina no interior das células quando submetidas ao estresse de 8 % e 10 % (v/v) de etanol. Este acúmulo acarretava em danos menores à parede celular e membrana plasmática pois mantinha a integridade das estruturas, além de promover a estabilização da morfologia e função das organelas devido a diminuição da geração de ROS. No ensaio pelo qPCR, foi observado a maior expressão de *ARG3* em 12 % de etanol (v/v), indicando que um maior acúmulo de arginina dentro de Y02999 pode estar contribuindo para a maior tolerância observada nesta linhagem nesta concentração de etanol.

O gene *BTN2* apresentou FC igual a 2,9377 e caracteriza-se por ser uma proteína de interação V-SNARE envolvida no tráfico intracelular de proteínas, que quando deletada causa erros de localização de diversas proteínas (Espinazo-Romeu *et al.*, 2008). Kim *et al.* (2005) demonstraram a interação de Btn2p com a proteína Rhb1p, responsável por regular a atividade da proteína de membrana Can1p, que é uma permease de arginina e lisina, ou seja, está envolvida na homeostase destes aminoácidos no interior da célula. A deleção de *BTN2* leva à alteração na captação destes aminoácidos.

Espinazo-Romeu *et al.* (2008) mostraram que a deleção de *BTN2* diminui a tolerância a etanol em linhagens industriais submetidas ao estresse em meio YPD acrescido de 15 % de etanol (v/v). Yamauchi *et al.* (2016) mostraram que os mRNAs de *BTN2* são eficientemente traduzidos em condições de estresse severo por etanol (9 % e

10 %) e que os níveis da proteína Btn2p diminuam após a eliminação do etanol. Compilando as informações obtidas de *BTN2*, hipotetiza-se que na presença do estresse por etanol este gene está mais expresso para ajudar a manter captação de aminoácidos que exercem efeitos protetores. Como Y02999 é uma linhagem mais tolerante, possivelmente a maior presença de *BTN2* possa estar contribuindo com esta característica. Pelo experimento de qPCR a diferença de expressão entre as duas linhagens não significativamente diferente, porém houve um perfil de maior expressão em Y02999 nas condições de 8 % e 12 % de etanol (v/v) validando o que foi observado no RNAseq e reforçando a hipótese acima.

- Genes *down*-regulados em Y02999

Entre os genes *down*-regulados descritos no Anexo A os que foram escolhidos para estudos seguintes são *ULI1*, *PEX17* e *COX23*.

*ULI1* é um gene induzido pela resposta de proteínas mal-dobradas (UPR) do retículo endoplasmático (Metzger *et al.*, 2009). Navarro-Tapia *et al.* (2017) demonstraram a indução das vias de resposta UPRs durante o estresse de 8 % de etanol (v/v) em *S. cerevisiae* BY4743 e ainda mostraram o impacto da tolerância ao etanol em linhagens deletadas para genes envolvidos na resposta UPR: *IRE1*, *HAC1* e *GCN4*. Quando estes três genes foram deletados, *S. cerevisiae* se mostrou significativamente menos tolerante em 8 % e 10 % de etanol (v/v) comparada com a linhagem parental. Analisando a expressão dos genes *IRE1*, *HAC1* e *GCN4* no presente trabalho, foi observado que eles não se apresentaram diferencialmente expressos comparando BY4741 e Y02999. Este resultado sugere que nem todos os genes envolvidos na via UPR tem sua expressão alterada na deleção de *APJ1*, mas isto não significa que a *down*-regulação observada para *ULI1* não esteja contribuindo para a maior tolerância ao etanol vista em Y02999.

O gene *PEX17* também é um gene *down*-regulado em Y02999, estando aproximadamente 4 vezes mais expresso em BY4741. Este gene codifica uma peroxina que faz parte do complexo de importação de proteínas componentes da matriz do peroxissomo (Brown *et al.*, 2003). Yoshikawa *et al.* (2009) relataram a importância dos genes envolvidos no transporte de proteínas peroxissomais em relação ao processo de tolerância em *S. cerevisiae* BY4742 quando submetida ao estresse de 8 % de etanol (v/v). Linhagens deletadas dos genes *PEX8*, *PEX14*, *PEX15*, *PEX17*, *PEX19* e *PEX22* se mostraram sensíveis à presença do etanol. Uma das funções do peroxissomo é a defesa contra o estresse oxidativo causado pela geração de ROS (Brown *et al.*, 2003), e sabe-se

que o estresse por etanol leva a um estresse oxidativo. Analisando a expressão dos outros genes responsáveis pela biogênese do peroxissomo descritos acima, não foi encontrado nenhum que estivesse diferencialmente expresso entre as linhagens, somente *PEX17*. Apesar do peroxissomo aparentemente apresentar um papel importante contra o estresse oxidativo, isto não quer de dizer que a *down*-regulação de *PEX17* não esteja contribuindo para o aumento da tolerância de Y02999, tendo em vista que a tolerância é um processo complexo.

*COX23* é um gene que também aparece *down*-regulado em Y02999, estando cerca 5,5 vezes mais expresso em BY4741. Este gene codifica para uma proteína solúvel localizada no espaço intermembrana da mitocôndria que é responsável pela homeostase de cobre na mitocôndria, além de ser essencial para expressão da citocromo *c* oxidase (complexo IV da cadeia de transporte de elétrons) (Barros *et al.*, 2004). Yoshikawa *et al.* (2008) mostraram que a deleção de genes envolvidos na biossíntese da mitocôndria, ubiquinona e citocromo *c* oxidase aumenta a sensibilidade de *S. cerevisiae* ao etanol por causar deficiências na respiração aeróbica. Ao avaliar a expressão de alguns outros genes que codificam proteínas que compõem o complexo IV, foi visto que a maioria não se encontra *down*-regulado em Y02999 como *COX23*, assim não é possível afirmar que o complexo IV não está funcional nesta linhagem. Hipotetiza-se que a baixa expressão desse gene possa ser o início de uma resposta de Y02999 ao estresse oxidativo causado pela geração de ROS na exposição ao etanol. Sabendo que genes relacionados a perda de DNA mitocondrial se encontram *up*-regulados, talvez seja o processo de perda de DNA mitocondrial o responsável pela diminuição das atividades da mitocôndria e desregulação de genes que codificam proteínas componentes dos complexos da cadeia de transporte de elétrons.

- Genes ausentes em Y02999

Entre os genes ausentes em Y02999 na análise pelo RNAseq, os genes *HES1*, *COS12* e *GMC2* se destacaram por estarem relacionados ao processo de tolerância e aparentemente não possuem ligação a maior tolerância de Y02999.

*HES1* é um gene envolvido na biossíntese de ergosterol (Jiang *et al.*, 1994). Em condições de estresse por etanol já foi relatado a importância da presença deste lipídeo para aumentar a rigidez da membrana plasmática e diminuir a fluidez causada por altas concentrações de etanol (Ding *et al.*, 2009). Qiu *et al.* (2017), ao fazer mutações aleatórias na subunidade Rpb7 da RNA polimerase II, conseguiu identificar um mutante de *S.*

*cerevisiae* BY4742 mais tolerante ao etanol comparado à linhagem controle, em meio YPD sólido acrescido de 10 % de etanol (v/v). Ao analisar os genes diferencialmente expressos no mutante por DNA *microarray* e qPCR, os genes *HES1*, *ERG4*, *ERG20* e *UPC2* envolvidos na síntese de ergosterol se encontravam *up*-regulados. Em uma análise de RNAseq feita por Li *et al.* (2017), submetendo a linhagem *S. cerevisiae* Sc131 ao estresse com 10 % de etanol (v/v) em meio YPD por 4 horas, foi visto que os genes *ERG2*, *ERG3* e *ERG24*, estavam *up*-regulados. No presente trabalho, todos os genes descritos envolvidos na síntese de ergosterol estavam expressos tanto na linhagem BY4741 e Y02999, mas não apresentaram diferenças de expressão e *HES1* não se encontrava presente em Y02999. A partir destas observações podemos pensar que os genes avaliados não são genes que explicam a maior tolerância ao etanol observada em Y02999.

*GMC2* é um gene específico de meiose envolvido na formação do complexo sinaptonêmico, responsável por aproximar os cromossomos homólogos para ocorrência do *crossing-over*. A formação deste complexo é mediada pela proteína Zip1p, que forma filamentos transversais responsáveis por juntar os cromossomos. Gmc2p aparentemente está envolvida com a sumoilação de Ecm11p, que é outra proteína relacionada à formação do complexo. Juntas, Gmc2p-Ecm11p, facilitam a montagem de Zip1p, sendo que a sumoilação de Ecm11p é crucial para ação de Zip1p (Humphries *et al.*, 2013).

Kasavi *et al.* (2016) demonstraram que a deleção do gene *PMT7*- que codifica para uma provável O-manosiltransferase possivelmente envolvida na manutenção da rigidez e estrutura da parede celular - em *S. cerevisiae* BY4743 leva ao aumento da tolerância ao etanol. Um estudo do transcriptoma desta linhagem deletada submetida a um estresse de 8 % de etanol (v/v) por duas horas, demonstrou que *GMC2* se encontrava *up*-regulada. Na análise pelo RNAseq deste trabalho foi observado a ausência de *GMC2* em Y02999 enquanto *ECM11* se encontrava 3,6798 vezes mais expressa nesta mesma linhagem e *PMT7* se encontrava um pouco menos presente ( $\log_2$  FC = -0,441). No experimento de qPCR em 8 % de etanol (v/v) para validação da ausência de expressão deste gene em Y02999, foi observado um comportamento de menor expressão em Y02999. Em 12 % de etanol (v/v) foi visto o comportamento contrário. Pode-se pensar que em 8 % de etanol (v/v) a *down*-regulação esteja respondendo a maior tolerância de Y02999 e com o aumento da concentração de etanol, a célula desenvolve novas maneiras de se adaptar à nova condição de estresse e a *up*-regulação de *GMC2* em Y02999 passa a ser uma forma de resposta da célula.

O gene *COS12* que codifica para uma proteína endossomal envolvida no *turnover* de proteínas de membrana não teve a expressão detectada em Y02999. Alper *et al.* (2006) mostraram que a dupla mutação para *SPT15* (que codifica uma proteína de ligação ao TATA box e é componente do fator de transcrição basal TFIID) e *COS12* em *S. cerevisiae* BY4741 implicou na diminuição do crescimento quando cultivadas na presença de 5 % de etanol (v/v) e 6 % de glicose (m/v), indicando a sensibilidade do duplo mutante ao etanol. No experimento de qPCR deste trabalho não foi possível ver diferença de expressão significativa entre as linhagens, porém foi visto um comportamento de *down-regulação* em Y02999 nas concentrações de 8 % e 12 % de etanol (v/v). Mesmo que Alper *et al.* (2006) tenha mostrado o impacto negativo da deleção de *COS12* no processo de tolerância, não seria surpresa se a *down-regulação* neste trabalho sugerisse a maior tolerância observada em Y02999, tendo em vista que o processo de tolerância é muito complexo.

- Genes presentes somente em Y02999

Dentre os 18 genes expressos somente em Y02999, encontram-se onze genes de funções ainda desconhecidas, três retrotransposons, dois pequenos RNAs nucleares, um gene envolvido na esporulação, responsável pela formação da membrana do esporo (*PFS1*) e *COS2*, que codifica para uma proteína endossomal envolvida no *turnover* de proteínas de membrana plasmática. Pouca informação pôde ser inferida do papel destes genes no efeito da deleção de *APJ1* na condição de estresse por etanol. Estudos posteriores dos genes desconhecidos são necessários para uma correta correlação entre o impacto da deleção na expressão destes genes e a tolerância ao etanol, além de uma futura validação por outra metodologia.

- Genes avaliados no qPCR que foram escolhidos com base na literatura

Dois genes foram escolhidos para serem avaliados pelo qPCR com base nos papéis descritos para eles na literatura em condições de estresse por etanol em *S. cerevisiae*. Eles são *OLE1* e *HSP104*.

O gene *OLE1* codifica uma  $\Delta^9$  dessaturase necessária para a mono-insaturação de ácidos graxos saturados, como ácido palmítico e ácido esteárico (Stukey *et al.*, 1990). Uma resposta clássica ao estresse por etanol em *S. cerevisiae* é a alteração da composição dos lipídeos da membrana plasmática para contornar a o aumento da fluidez causado pelo etanol e estabilizar a membrana (Stanley *et al.*, 2010). Essa alteração correspondente ao

aumento de ácidos graxos insaturados e ergosterol. As concentrações de ácido palmitoléico e o ácido oléico são maiores em linhagens mais tolerantes ao etanol e a insaturação de ambos os lipídeos foi gerada pela atividade da dessaturase codificado pelo gene *OLE1* (Archana *et al.*, 2015). You *et al.* (2003) e Nasution *et al.* (2016) mostraram que a superexpressão de *OLE1* em linhagens de *S. cerevisiae* laboratoriais aumentou a tolerância das mesmas quando submetidas ao estresse de 5 % de etanol (v/v). Na análise de qPCR deste trabalho, em 8 % de etanol (v/v) o gene *OLE1* não estava diferencialmente expresso entre as linhagens, indicando que nessa concentração de etanol a deleção de *APJ1* não altera a expressão deste gene. Na condição de 12 % de etanol (v/v), foi observado a *down*-regulação de *OLE1* em Y02999, significando que a diminuição da expressão deste gene possa estar contribuindo para a maior tolerância de Y02999 nesta concentração de etanol. Apesar da literatura mostrar a importância da presença de *OLE1* no aumento da tolerância, isto não significa que sua menor expressão em Y02999 não esteja contribuindo positivamente para a tolerância.

O gene *HSP104* é uma desagregase da classe das chaperonas Hsp100p que possui um domínio AAA+ ATPase conservado. Cooperar com as atividades das Hsp70p e Hsp40p no desdobramento e reativação de proteínas agregadas (Yokom *et al.*, 2016). A associação com as pequenas Hsps facilita a ação das Hsp104p no reparo e redobramento das proteínas (Lytras *et al.*, 2017). Esta chaperona já foi descrita como *up*-regulada em condições de estresse por etanol (Ding *et al.*, 2009; Stanley *et al.*, 2010). O gene *HSP104* teve sua expressão avaliada pelo qPCR e foi visto que em 12 % de etanol (v/v) ele se encontra *up*-regulado em Y02999, podendo sugerir que a maior tolerância desta linhagem nesta concentração de etanol está relacionada com a maior resposta as proteínas mal dobradas que surgem no estresse. Apj1p é uma putativa chaperona da família das Hsp40p contendo um domínio DNA J conservado. Este grupo de chaperonas é responsável por regular a atividade ATPase das Hsp70p (Fan *et al.*, 2003). Hsp70p compreende chaperonas envolvidas na prevenção da agregação proteica, dobramento correto das proteínas e solubilização e redobramento de proteínas agregadas, sendo que todas estas funções são guiadas pela ligação e hidrólise de ATP (Vogel *et al.*, 2006). O mecanismo pela qual a deleção de uma chaperona (Apj1p) leva ao aumento da tolerância ao etanol em leveduras ainda não é conhecido, mas a *up*-regulação de *HSP104* sugere que a deleção estimula a produção de outras Hsps.

Outros dois genes escolhidos para o qPCR foram *UBC9* e *SIZ1*. Estes genes são componentes da via de sumoilação, (modificação pós-traducional semelhante à

ubiquitinação), sendo Ubc9p a enzima de conjugação E2, que contém um resíduo de cisteína onde o sumo é ligado, e Siz1p correspondente a ligases E3, que estimula a transferência do SUMO a proteína que receberá a modificação (Enserink, 2015).

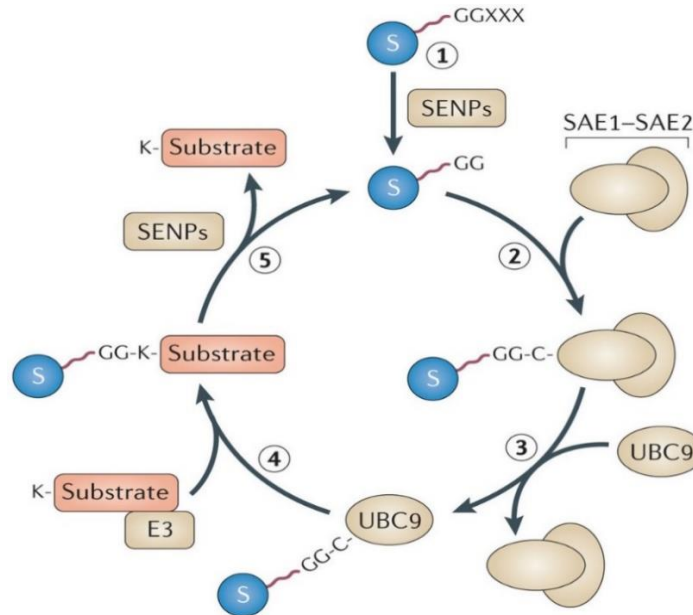


Figura 24: Sumoilação de proteínas alvo. Na etapa 1, ocorre a maturação do SUMO, onde proteases específicas (SENPs) expõem um motivo glicina-glicina. Na etapa 2, o SUMO maduro é ligado a um resíduo de cisteína na enzima E1 (heterodímero SAE1-SAE2). Nas etapas 3 e 4, o SUMO é transferido para um resíduo de cisteína da enzima de conjugação E2 (Ubc9). Esta enzima catalisa a conjugação do SUMO a um resíduo de lisina da proteína alvo, de forma independente ou juntamente com as E3 ligases. Fonte: Everett *et al.* (2013).

O interesse nestes genes veio do envolvimento de *APJ1* na degradação de proteínas sumoiladas (Sahi *et al.*, 2013). Zhou *et al.* (2004) relataram que na presença de 10 % de etanol (v/v) e 1 mM de H<sub>2</sub>O<sub>2</sub> ocorre um grande aumento de proteínas com esta modificação. Dentre as proteínas sumoiladas descritas por Zhou *et al.* (2004), estavam as próprias proteínas do complexo de sumoilação (indicando auto-sumoilação), proteínas envolvidas em repostas de danos ao DNA, algumas Hsps, Sod1p responsável pela eliminação de ROS, dentre outras.

A ideia da investigação destes genes foi entender transcricionalmente, como os genes componentes da maquinaria de sumoilação se comportariam na ausência de *APJ1*. Esperava-se que eles se apresentassem *down*-regulados em Y02999, tendo em vista que *Apj1p* não estaria presente na célula para degradar as proteínas com o sumo e assim não seria necessária uma maior transcrição dos genes da maquinaria para repor as proteínas

sumoiladas dentro da célula. Estas proteínas sumoiladas presentes no estresse por etanol aparentemente contribuem positivamente a tolerância das leveduras. A expressão de *UBC9* e *SIZ1* foi avaliada por qPCR e não foi obtido diferenças significativas de expressão entre BY4741 e Y02999 nas condições de 8 % e 12 % de etanol. Com estes resultados, é possível concluir que a expressão dos genes da maquinaria de sumoilação não é controlada em nível transcricional e possivelmente existe um controle traducional ou pós-traducional responsável por manter a homeostase de proteínas sumoiladas no interior da célula na presença do estresse e ausência de Apj1p.

## 6.2 CRISPR-Cas9 utilizando o vetor pML104 em *K. marxianus* CCT7735

*Kluyveromyces marxianus* é uma levedura que apresenta características favoráveis para aplicações industriais, como fermentação à altas temperaturas (> 40 °C), rápido crescimento, capacidade de utilização de diversas fontes de carbono para produção de etanol e fraca repressão por glicose, ou seja, não é necessário o esgotamento da glicose para que outras fontes de carbono sejam fermentadas. Sua capacidade de tolerar altas temperaturas e crescer em diversos substratos a torna interessante para o processo de sacarificação e fermentação simultâneos, o que diminuiria os custos de resfriamento e os riscos de contaminação do processo de produção de etanol. Porém, esta levedura possui algumas limitações, como o aumento da respiração na presença de pentoses, o que resulta em uma baixa produção de etanol. Também apresenta baixa tolerância a altas concentrações de etanol, o que limita sua utilização industrial para produção deste biocombustível (Gao *et al.*, 2015; Lertwattanasakul *et al.*, 2013; Lertwattanasakul *et al.*, 2015; Diniz *et al.*, 2017). Costa *et al.* (2014), mostraram que a *K. marxianus* CCT7735 (linhagem utilizada neste trabalho), é capaz de crescer em meio YPD com apenas 6 % de etanol (v/v) quando incubada a 30 °C, nenhum crescimento foi observado em 8 % e 10 % de etanol (v/v). Mostraram também que o aumento da temperatura diminui ainda mais a tolerância, a 45 °C esta linhagem cresceu somente em 2 % de etanol (v/v), não apresentando nenhum crescimento em 4 %, 6 %, 8 % e 10 % (v/v).

Como a deleção de *APJ1* mostrou um fenótipo de aumento de tolerância em *S. cerevisiae*, surgiu a ideia da utilização do sistema CRISPR-Cas9 para silenciamento deste gene em *K. marxianus* CCT7735, na tentativa de promover o aumento da tolerância ao etanol. O sistema escolhido foi o vetor pML104, que expressa tanto a Cas9 como o gRNA. O funcionamento deste vetor foi validado por Laughery *et al.* (2015) em *S. cerevisiae*, mostrando que quando a levedura era transformada com o pML104 clonado com o gRNA

a porcentagem de transformantes mutantes para o gene de interesse foi 23 % (de 13 colônias investigadas, 3 apresentavam a mutação). Essa porcentagem aumentou para 97 % (de 36 colônias investigadas) quando juntamente com o vetor, foi transformado um híbrido de oligos de 90 nucleotídeos que recombina homologamente dentro do gene alvo. Estes oligos apresentavam uma mutação na sequência 5'-NGG e ao ser inserido por recombinação, alterava o quadro de leitura dos códons e a proteína final codificada pelo gene perdia funcionalidade.

A edição pelo sistema CRISPR-Cas9 pode ocorrer pelos dois sistemas de reparo de quebra de dupla fita, a junção de extremidades não homólogas (NHEJ) e o reparo por recombinação homóloga (RH). *K. marxianus*, diferentemente de *S. cerevisiae*, possui o mecanismo NHEJ muito ativo e usa pouco o RH. Sendo assim, a taxa de edição com um DNA doador é baixa pois ele não integra eficientemente no gene alvo, fazendo integrações aleatórias no genoma (Juergens *et al.*, 2018; Nambu-Nishida *et al.*, 2017). Por isso neste trabalho não foi desenhado um DNA doador e a edição deveria ocorrer pelo reparo NHEJ.

A clonagem do gRNA não ocorreu da forma correta, inserindo duas cópias do gRNA ao invés de uma única cópia. A figura 25 ilustra a clonagem esperada dentro do pML104 e a clonagem com os dois guias obtida. Apesar disso, o gRNA 2 indicado na figura 25-B permaneceu na orientação correta e ao ser transcrito vai gerar um gRNA que possivelmente pode guiar a clivagem pela Cas9. Além disso, ao gRNA 2 será adicionado corretamente a extensão do gRNA presente no pML104. Ao ser transcrita, essa extensão gera uma estrutura de RNA secundária necessária para o reconhecimento do ponto de clivagem pela Cas9.



Figura 25: **A.** Clonagem esperada de somente um gRNA dentro do pML104. **B.** Clonagem obtida com dois gRNAs (1 e 2) dentro do vetor.

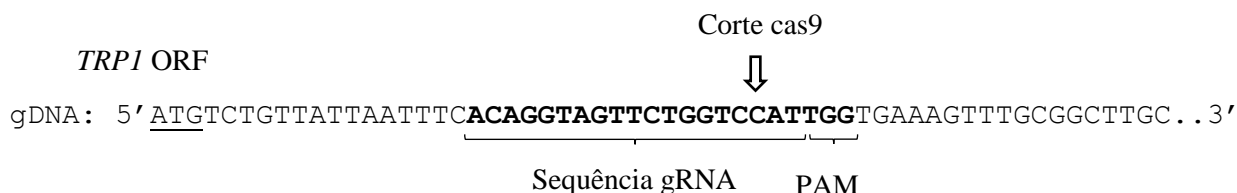
O experimento para verificar a expressão de Cas9 em *S. cerevisiae* e *K. marxianus* foi inconclusivo, já que os dois *primers* desenhados para o experimento anelavam de forma inespecífica no genoma das leveduras não transformadas. Esperava-se, mesmo com a inespecificidade, que nas leveduras transformadas a banda observada no gel fosse mais forte, indicando que além da banda inespecífica haveria a presença de cDNA de Cas9. Este perfil não foi observado, levando a pensar que a Cas9 não estaria sendo expressa em nenhuma das duas leveduras.

Uma questão levantada para utilização do vetor em *K. marxianus* era se os promotores que controlam a expressão do gRNA e Cas9 iriam ser reconhecidos pela maquinaria de transcrição da célula. O promotor para o gRNA é o pSNR52 de *S. cerevisiae*, que é um promotor reconhecido pela RNA polimerase III, e para Cas9 é o pTDH3 de *S. cerevisiae* que é um promotor forte reconhecido pela RNA polimerase II, (Juergens *et al.*, 2018). Porém, Juergens *et al.* (2018), mostraram a utilização do pTDH3 de *S. cerevisiae* para expressão do gRNA em *K. marxianus* NBRC1777 e *K. marxianus* CBS397 e a edição funcionou nas duas linhagens. Löbs *et al.* (2017) utilizaram o promotor pSNR52 de *S. cerevisiae* para expressar o gRNA em *K. marxianus* 6556 e também tiveram sucesso na edição, apesar de Juergens *et al.* (2018) ter relatado que este promotor talvez não possa ser utilizado por todas as espécies de leveduras. Estes trabalhos nos levaram a concluir que a Cas9 possivelmente está sendo expressa pela maquinaria de transcrição, porém ela pode não estar sendo traduzida de forma eficiente ou necessita de uma otimização dos códons para aumentar a eficiência de edição em *K. marxianus*, como feito por Löbs *et al.* (2017). Outra questão era se Cas9 poderia ser tóxica para célula. Stovicek *et al.* (2017), relataram trabalhos em que a presença de Cas9 causava limitações no crescimento em leveduras quando a expressão era controlada por promotores fortes. Porém, este não foi o problema deste trabalho, pois muitos transformantes cresceram na placa com meio seletivo e quando inoculados em meio seletivo líquido cresciam bem em 16 horas.

O sequenciamento de Sanger de 4 colônias selecionadas aleatoriamente, mostrou que não houve nenhuma alteração de nucleotídeo na região de corte pela Cas9. A figura 26-A mostra o ponto de corte da Cas9 esperado por Laughery *et al.* (2015). Foi nesta mesma região que a alteração de bases foi investigada no presente trabalho, figura 26-B, tendo em vista que o gRNA foi desenhado para um 5' TGG que se encontrava no começo da sequência de *APJI*, próximo ao ATG inicial. A escolha deste gRNA foi estratégica,

pois quando a alteração ocorre no início da região codificadora maiores são as chances de truncar a proteína e fazer com que ela perca suas funções biológicas.

**A**



**B**

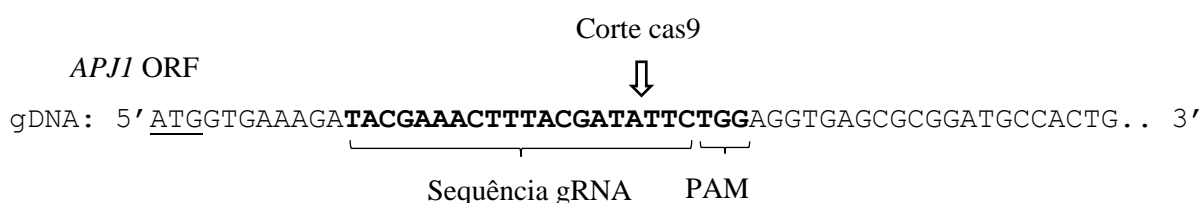


Figura 26: **A.** Esquema do 5' NGG (sequência PAM) escolhido para guiar a clivagem da Cas9 no gene *TRP1* de *S. cerevisiae*. O gRNA corresponde à Sequência gRNA indicada na figura e o ATG sublinhado corresponde ao ATG de início da tradução, Fonte: Modificado de Laughery *et al.* (2015). **B.** Esquema do 5' NGG escolhido para guiar a clivagem em *APJI* de *K. marxianus* CCT7735.

O número de colônias investigadas por sequenciamento foi pouco tendo em vista que a placa resultante da transformação estava toda preenchida com colônias. Sabendo que a porcentagem de colônias editadas encontrada por Laughery *et al.* (2015) foi baixa quando a edição ocorreu pelo reparo NHEJ, talvez seja interessante fazer futuramente uma amostragem maior de colônias para investigação e comprovação do não funcionamento do sistema. Novas tentativas de digestão e clonagem também devem ser realizadas para clonar somente um guia no vetor e aumentar as chances do RNA guia anelar corretamente em seu alvo.

Outras alternativas de sistemas para edição também são possíveis. Várias estratégias de construção de vetores têm sido feitas para aplicar o sistema CRISPR-Cas9 nas mais diversas espécies de leveduras, como por exemplo, *K. marxianus*, *K. lactis*, *Ogataea polymyrpha*, *Candida glabrata* e *Yarrowia lipolytica* (Enkler *et al.*, 2016; Juergens *et al.*, 2018; Schwartz *et al.*, 2015). No caso de *K. marxianus*, o bloqueio de genes envolvidos no sistema de reparo NHEJ tem funcionado como estimulador do reparo

por recombinação, como mostrado por Nambu-Nishida *et al.* (2017). Neste trabalho, os genes *NEJ1* e *DNL4*, envolvidos no mecanismo de reparo NHEJ, foram inativados por mutação pontual e o mutante para cada gene foi transformado com um oligo que possuía 50 pares de bases de homologia com o gene alvo e recombinava no ponto de clivagem da Cas9. Foi visto para a linhagem *wild-type* uma taxa de 28 % de integração dentro do alvo, sendo que para o mutante *NEJ1* esta taxa foi de 84 % e para o mutante *DNL4* a taxa foi de 92 %, indicando o aumento da eficiência do reparo por homologia.

## 7. CONCLUSÃO

Pela análise comparativa do transcriptoma de *S. cerevisiae* BY4741 e Y02999 e avaliação de alguns genes candidatos por qPCR, foi possível concluir que a deleção de *APJI* altera a expressão de genes na presença do estresse de 12 % de etanol (v/v), como *HSP104*, *ARG3* e *OLE1*. Estes genes diferencialmente expressos estão relacionados a maior tolerância observada em Y02999. Também foi possível concluir, que a transcrição de genes componentes da maquinaria de sumoilação das células (*UBC9* e *SIZ1*) não é alterada na presença do etanol, significando que o controle da sumoilação na presença do etanol não é transcricional. Como perspectiva para este trabalho, tem-se a validação da expressão por qPCR de outros genes diferencialmente expressos entre as linhagens na análise do transcriptoma.

Para a edição de *APJI* em *K. marxianus* CCT7735 utilizando o vetor pML104, não foi possível concluir a efetividade do sistema pois poucas colônias foram investigadas. Como perspectiva, fica a investigação de um número maior de transformantes. Também pode ser feito o bloqueio de genes envolvidos no reparo NHEJ afim de estimular a recombinação homóloga e aumentar as chances de sucesso da edição.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ADITIYA, H, B. Second generation bioethanol production. A critical review. **Renewable and Sustainable Energy Reviews**, v. 66, p. 631-653, 2016.

AGUILERA, F. *et al.* Relationship between ethanol tolerance, H<sup>+</sup>-ATPase activity and the lipid composition of the plasma membrane in different wine yeast strains. **International Journal of Food Microbiology**, v. 110, p. 34-42, 2006.

ALEXANDRE, H. *et al.* Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*. **FEBS Letters**, v. 498, p. 98-103, 2001.

ALMEIDA, J. R. M. *et al.* Stress-related challenges in pentose fermentation to ethanol by the yeast *Saccharomyces cerevisiae*. **Biotechnology Journal**, v. 6, p. 286-299, 2011.

ALPER, H. *et al.* Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production. **Science**, v. 314, p. 1565-1596, 2006.

ALVIRA, P. *et al.* Pretreatment Technologies for efficient bioethanol production process base on enzymatic hydrolysis: A review. **Bioresource Technology**, v. 101, p. 4851-4861, 2010.

AMORIM, H. V. *et al.* Scientific challenges of bioethanol production in Brazil. **Applied Microbiology and Biotechnology**, v. 91, p. 1267-1275, 2011.

ANDREWS, S. FastQC: a quality control tool for high throughput sequence data. Disponível online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, 2010.

ARCHANA, K. *et al.* Correlation between ethanol stress and cellular fatty acid composition of alcohol producing non-*Saccharomyces* in comparison with *Saccharomyces cerevisiae* by multivariate techniques. **Journal of Food Science Technology**, v. 52, p. 6770-6776, 2015.

AUESUKAREE, C. Molecular Mechanisms of the yeast adaptive response and tolerance to stresses encountered during ethanol fermentation. **Journal of Bioscience and Bioengineering**, v.124, p. 133-142, 2017.

BELO, E. V. C. Construção de uma linhagem flocculante de *Kluyveromyces marxianus* com potencial para produzir etanol. Dissertação de mestrado apresentada à Universidade Federal de Viçosa, 2013.

BOLGER, A. M. *et al.* Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, p. 2114-2120, 2014.

BROWN, L. *et al.* Peroxisome biogenesis and the role of protein import. **J. Cell. Mol. Med.**, v. 7, p. 388-400, 2003.

CASTRO, A. M., PEREIRA JR., N. Produção, propriedades e aplicação de celulases na hidrólise de resíduos agroindustriais. **Química Nova**, v. 33, p.181-188, 2010.

CASPETA, L. *et al.* Modifying Yeast Tolerance to inhibitory Conditions of ethanol Production Processes. **Frontiers in Bioengineering and Biotechnology**, v. 3, 2015.

CHAROENBHAKDI, S. *et al.* Vacuolar H<sup>+</sup>-ATPase protects *Saccharomyces cerevisiae* cells against ethanol-induced oxidative and cell wall stresses. **Applied and Environmental Microbiology**, v. 82, p. 3121-3130, 2016.

- CHEN, J. *et al.* A convenient Cas9-based Conditional Knockout Strategy for Simultaneously Targeting Multiple Genes in Mouse. **Scientific Reports**, p. 1-7, 2017.
- CHENG, Y. *et al.* Protective effects of arginine on *Saccharomyces cerevisiae* against ethanol stress. **Sci. Rep.**, v. 6, p. 31311-31323, 2016.
- CHU, Y., COREY, D. R. RNA Sequencing: Platform Selection Experimental Design, and Data Interpretation. **Nucleic Acid Therapeutics**, v. 22, p. 271-274, 2012.
- COLLART, M. A., OLIVIERO, S. Preparation of Yeast RNA. **Current Protocols in Molecular Biology**, chapter 13, unit 13.12, 2001.
- CONESA, A. *et al.* A survey of best practices for RNA-seq data analysis. **Genome Biology**, p. 1-19, 2016.
- COSTA, D. *et al.* Physiological characterization of thermotolerant yeast for cellulosic ethanol production. **Applied Microbiology and Biotechnology**, v. 98, p. 3829-3840, 2014.
- DEMEKE, M. M. *et al.* Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. **Biotechnology for Biofuels**, v. 6, p. 89-103, 2013.
- DiCARLO, J. E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. **Nucleic Acids Research**, v. 41, p. 4336-4343, 2013.
- DING, J. *et al.* Tolerance and stress response to ethanol in the yeast *Saccharomyces cerevisiae*. **Applied Microbiology and Biotechnology**, v. 85, p. 253-263, 2009.
- DINIZ, R. *et al.* Transcriptome analysis of the thermotolerant yeast *Kluyveromyces marxianus* CCT 7735 under ethanol stress. **Applied Microbiology and Biotechnology**, v. 101, p. 6969-6980, 2017.
- DUROUX-RICHARD, I. *et al.* CRISPR-Cas9: A revolution in genome editing in rheumatic diseases. **Joint Bone Spine**, v.84, p. 1-4, 2017.
- EL-SAYED, A. S. A. *et al.* Genome editing approaches: manipulating of lovastatin and taxol synthesis of filamentous fungi by CRISPR/Cas9 system. **Applied Microbiology and Biotechnology**, v. 101, p. 3953-3976, 2017.
- ENKLER, L. *et al.* Genome engineering in the yeast pathogen *Candida glabrata* using the CRISPR-Cas9 system. **Sci. Rep.**, v. 6, p. 35766-35777, 2016.
- ENSERINK, J. M. Sumo and the cellular stress response. **Cell Division**, v.10, p. 1-13, 2015.
- ESPINAZO-ROMEU, M. *et al.* Btn2p is involved in ethanol tolerance and biofilm formation in flor yeast. **FEMS Yeast Res**, v. 8, p. 1127-1136, 2008.
- EVERETT, R.D. *et al.* Interplay between viruses and host sumoylation pathways. **Nature Reviews Microbiology**, v.11, p. 400-411, 2013.
- FAN, C. *et al.* Mechanisms for regulation of Hsp70 function by Hsp40. **Cell Stress & Chaperones**, v. 8, p. 309-316, 2003.

- FENG, X., ZHAO, H. Investigating host dependence of xylose utilization in recombinant *Saccharomyces cerevisiae* strains using RNA-seq analysis. **Biotechnology for biofuels**, p. 1-17, 2013.
- GAO, J. *et al.* Transcriptional analysis of *Kluyveromyces marxianus* for ethanol production from inulin using consolidated bioprocessing technology. **Biotechnol Biofuels**, v. 8, p. 115-132, 2015.
- GIETZ, R. D. *et al.* High- efficiency yeast transformation using the LiAc/ SS carrier DNA/ PEG method. **Nature Protocol**, v. 2, p. 34-34, 2007.
- GOODWIN, S. *et al.* Coming of age: ten years of next-generation sequencing technologies. **Nature Reviews**, v. 17, p. 333-351, 2016.
- HALIM, D. *et al.* Loss of LMOD1 impairs smooth muscle cytocontractility and causes megacystis microcolon intestinal hypoperistalsis syndrome in humans and mice. **PNAS**, v.114, p. 2739-2747, 2017.
- HARNER, N. K. *et al.* Genetic improvement of native xylose-fermenting yeasts for ethanol production. **Journal of Industrial Microbiology and Biotechnology**, v. 42, p. 1-20, 2015.
- HESS, D. *et al.* Computationally Driven, Quantitative Experiments Discover Genes Required for Mitochondrial Biogenesis. **PLoS Genetics**, v. 5, p. e1000407-e1000423, 2009.
- HIRASAWA, T. *et al.* Identification of target genes conferring ethanol stress tolerance to *Saccharomyces cerevisiae* based on DNA microarray data analysis. **Journal of Biotechnology**, v. 131, p. 33-44, 2007.
- HOU, X. *et al.* Improved inhibitor tolerance in xylose-fermenting yeast *Spathaspora passalidarum* by mutagenesis and protoplast fusion. **Applied Microbiology and Biotechnology**, v. 93, p. 2591-2601, 2012.
- HUMPHRYES, N. *et al.* The Ecm11-Gmc2 complex promotes synaptonemal complex formation through assembly of transverse filaments in budding yeast. **PLoS Genetics**, v. 9, p. e1003194 – e1003208, 2013.
- JIANG, B. *et al.* A new family of yeast genes implicated in ergosterol synthesis is related to the human oxysterol binding protein. **Yeast**, v.3, p. 341- 353, 1994.
- JUERGENS, H. *et al.* Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. Manuscript submetido à **FEMS Yeast Research**, 2017.
- JUNQUEIRA, T. L. *et al.* Techno-economic analysis and climate change impacts of sugarcane biorefineries considering different time horizons. **Biotechnology for biofuels**, p. 10-50, 2017.
- KANG, Q. *et al.* Bioethanol from Lignocellulosic Biomass: Current Findings Determine Research Priorities. **The Scientific World Journal**, p. 1-13, 2014.
- KASAVI, C. *et al.* An integrative analysis of transcriptomic response of ethanol tolerant strains to ethanol in *Saccharomyces cerevisiae*. **Mol BioSyst.**, v. 12, p. 464-477, 2016.

KIM, Y. *et al.* Interaction among Btn1p, Btn2p, and Ist2p reveals potential interplay among the vacuole, amino acid levels, and ion homeostasis in the yeast *Saccharomyces cerevisiae*. **Eukaryotic Cell**, v. 4, p. 281-288, 2005.

KUKURBA, K. R., MONTGOMERY, S. B. RNA Sequencing and Analysis. **Cold Spring Harbor Protocols**, v. 11, p. 951-969, 2015.

LANGMEAD, B. *et al.* Fast gapped-read alignment with bowtie 2. **Nature Methods**, v. 9, p. 357-360, 2012.

LAUGHERY, M. F. *et al.* New vectors for simple and streamlined CRISPR–Cas9 genome editing in *Saccharomyces cerevisiae*. **Yeast**, v. 32, p. 711-720, 2015.

LEE, Y. *et al.* Bioethanol production from cellulosic hydrolysates by engineered industrial *Saccharomyces cerevisiae*. **Bioresource Technology**, v. 228, p. 355-361, 2017.

LENNARTSSON, P. R. *et al.* Integration of the first and second generation bioethanol processes and the importance of by-products. **Bioresource Technology**, v. 165, p. 3-8, 2014.

LERTWATTANASAKUL, N. *et al.* Essentiality of respiratory activity for pentose utilization in thermotolerant yeast *Kluyveromyces marxianus* DMKU 3-1042. **Antonie van Leeuwenhoek**, v. 103, p. 933-945, 2013.

LERTWATTANASAKUL, N. *et al.* Genetic basis of the highly efficient yeast *Kluyveromyces marxianus*: complete genome sequence and transcriptome analyses. **Biotechnology for Biofuels**, v. 8, p. 47-61, 2015.

LI, R. *et al.* Investigating the underlying mechanism of *Saccharomyces cerevisiae* in response to ethanol stress employing RNA-seq analysis. **World J. Microbiol and Biotechnol**, v. 33, p. 206-2019, 2017.

LIANG, M. *et al.* Impact of pseudo-continuous fermentation on the ethanol tolerance of *Scheffersomyces stipitis*. **Journal of Bioscience and Bioengineering**, v. 116, p. 319-326, 2013.

LITVINCHUK, A. *et al.* Mitochondrially-encoded protein Var1 promotes loss of respiratory function in *Saccharomyces cerevisiae* under stressful conditions. **European Journal of Cell Biology**, v. 92, p. 163-174, 2013.

LÖBS, A. *et al.* CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in *Kluyveromyces marxianus*. **Biotechnology for Biofuels**, v. 10, p. 164-178, 2017.

LÖÖKE, M. *et al.* Extraction of Genomic DNA from yeasts for PCR based Applications. **Biotechniques**, p. 325-328, 2011.

LÜ, F. *et al.* Precise pretreatment of lignocellulose: relating substrate modification with subsequent hydrolysis and fermentation to products and by-products. **Biotechnology for biofuels**, v. 10, p. 88-103, 2017.

LYTRAS, G. *et al.* Asymmetric inheritance of the yeast chaperone Hsp26p and its functional consequences. **Biochemical and Biophysical Research Communications**, v. 491, p. 1055-1061, 2017.

- MA, M. *et al.* Quantitative transcription dynamic analysis reveals candidate genes and key regulators for ethanol tolerance in *Saccharomyces cerevisiae*. **BMC Microbiology**, 2010.
- MCRELI, S. *et al.* Techno-economic evaluation of 2nd generation bioethanol production from sugar cane bagasse and leaves integrated with the sugar-based ethanol process. **Biotechnology for biofuels**, v. 5, p. 22-40, 2012.
- McGETTIGAN, P. A. Transcriptomics in the RNA-seq era. **Current Opinion in Chemical Biology**, v. 17, p. 4-11, 2013.
- MEENA, R. *et al.* Regulation of *Saccharomyces cerevisiae* plasma membrane H<sup>+</sup>-ATPase (pma1) by dextrose and hsp30 during exposure to thermal stress. **Indian Journal of Microbiology**, v. 51, p. 153-158, 2011.
- MEI, Y. *et al.* Recent Progress in CRISPR/Cas9 Technology. **Journal of Genetics and Genomics**, v. 43, p. 63-75, 2016.
- MERCX, S. *et al.* Inactivation of the  $\beta(1,2)$ -xylosyltransferase and  $\alpha(1,3)$ -fucosyltransferase genes in *Nicotia tabacum* BY-2 Cells by Multiplex CRISPR/Cas9 Strategy Results in Glycoproteins without Plant-Specific Glycans. **Front. Plant Sci**, v. 8, p. 1-11, 2017.
- METZGER, M. *et al.* Analysis of quality control substrates in distinct cellular compartments reveals a unique role for Rpn4p in tolerating misfolded membrane proteins. **Molecular Biology of the Cell**, v. 20, p. 1006-1019, 2009.
- MILNE, I. *et al.* Tablet- Next generation sequence assembly visualization. **Bioinformatics**, v. 26, p. 401-402, 2010.
- NAMBU-NISHIDA, Y *et al.* Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. **Sci-rep**, v. 7, p. 8993-9001, 2017.
- NASUTION, O. *et al.* Overexpression of OLE1 enhances stress tolerance and constitutively activates the MAPK HOG pathway in *Saccharomyces cerevisiae*. **Biotechnology and Bioengineering**, v. 114, p. 620-631, 2016.
- NAVARRO-TAPIA, E. *et al.* Ethanol effects involve non-canonical unfolded protein response activation in yeast cells. **Frontiers in Microbiology**, v. 8, 2017.
- NOVA CANA<sup>1</sup>. Benefícios em usar etanol. Disponível em <https://www.novacana.com/etanol/beneficios/#economia-brasileira>. Acesso em 10 abr. 2017.
- NOVA CANA<sup>2</sup>. Matérias-primas do etanol de 2<sup>a</sup> geração. Disponível em <https://www.novacana.com/etanol/materias-primas>. Acesso em 10 abr. 2017.
- OGEDA, T. L., PETRI, D. F. S. Hidrólise Enzimática de Biomassa. **Química Nova**, São Paulo, v. 33, p.1549-1558, 2010.
- PAIS, T. M. *et al.* Comparative Polygenic Analysis of Maximal Ethanol Accumulation Capacity and Tolerance to High Ethanol Levels of Cell Proliferation in Yeast. **PLOS Genetics**, v. 9, 2013.

- PITARELO, A. P. *et al.* Efeito do teor de umidade sobre o pré-tratamento a vapor e a hidrólise enzimática do bagaço de cana-de-açúcar. **Química Nova**, Curitiba, p.1-8, 2012.
- QIU, Z. *et al.* Improving *Saccharomyces cerevisiae* ethanol production and tolerance via RNA polymerase II subunit Rpb7. **Biotechnology for biofuels**, v. 10, p. 125-138, 2017.
- RADECKA, D. *et al.* Looking beyond *Saccharomyces*: the potencial of non-conventional yeast species for desirable traits in bioethanol fermentation. **FEMS Yeast Research**, v.15, 2015.
- RAÍZEN<sup>1</sup>. Etanol de primeira geração. Disponível em <http://www.raizen.com.br/energia-do-futuro-tecnologia-em-energia-renovavel/etanol-de-primeira-geracao>. Acesso em 10 abr. 2017.
- RAÍZEN<sup>2</sup>. Etanol de segunda geração. Disponível em <http://www.raizen.com.br/energia-do-futuro-tecnologia-em-energia-renovavel/etanol-de-segunda-geracao>. Acesso em 10 abr. 2017.
- RAN, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. **Nature Protocols**, v. 8, p. 2281-2308, 2013.
- SAHI, C. *et al.* Sequential Duplications of an Ancient Member of the DNAJ-Family Expanded the Functional Chaperone Network in the Eukaryotic Cytosol. **Molecular Biology and Evolution**, v. 30, p. 985-998, 2013.
- SAMBROOK, J. *et al.* Molecular Cloning: A Laboratory Manual, **Cold Spring Harbor: Cold Spring Harbor Laboratory Press**, Plainview, 2ª edição, p. 49-55, 1989.
- SANDER, J. D.; JOUNG, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. **Nature Biotechnology**, v. 32, p. 347-355, 2014.
- SANTOS, J. R. A. *et al.* Comparação entre processos em SHF e em SSF de bagaço de cana-de-açúcar para a produção de etanol por *Saccharomyces cerevisiae*. **Química Nova**, v. 33, p. 904-908, 2009.
- SARKAR, N. *et al.* Bioethanol production from agricultural wastes: An overview. **Renewable Energy**, p.19-27, 2012.
- SCHWARTZ, C. *et al.* Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. **Synthetic Biology**, v. 5, p. 356-359, 2016.
- SENATHAM, S. *et al.* Enhanced xylose fermentation and hydrolysate inhibitor tolerance of *Scheffersomyces shehatae* for efficient ethanol production from non-detoxified lignocellulosic hydrolysate. **Springer Plus**, p. 1-8, 2016.
- SHALLOM, D., SHOHAM, Y. Microbial hemicelulases. **Current Opinion in Microbiology**, v. 6, p. 219-228, 2003.
- SNOEK, T. *et al.* How do yeast cells become tolerant to high ethanol concentrations? **Current Genetics**, v. 62, p. 475-480, 2016.

- SOUZA, W. R. *et al.* Transcriptome analysis of *Aspergillus niger* grown on sugarcane bagasse. **Biotechnology for biofuels**, p. 1-16, 2011.
- STANLEY, D. *et al.* The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. **Journal of Applied Microbiology**, v. 109, p. 13-24, 2010.
- STOVICEK, V. *et al.* CRISPR/Cas system for yeast genome engineering: advances and applications. **FEMS Yeast Research**, v. 17, p. 5 – 31, 2017.
- STUKEY, J. *et al.* The *OLE1* gene of *Saccharomyces cerevisiae* encodes the A9 fatty acid desaturase and can be functionally replaced by the rat stearyl-CoA desaturase gene. **The Journal of Biological Chemistry**, v. 265, p. 20144-20149, 1990.
- SURESH, T. *et al.* *Saccharomyces cerevisiae* Hsp30 is necessary for homeostasis of a set of thermal stress response functions. **Journal of Microbiology and Biotechnology**, v. 20, p. 403-409, 2010.
- SWINNEN, S. *et al.* Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. **Genome Research**, v. 22, p. 975-984, 2012.
- SYAHIDAK, K. *et al.* Time-based comparative transcriptomics in engineered xylose-utilizing *Saccharomyces cerevisiae* identifies temperature-responsive genes during ethanol production. **Journal of Industrial Microbiology and Biotechnology**, v. 40, p. 1039-1050, 2013.
- TAKAGI, H. *et al.* Proline accumulation protects *Saccharomyces cerevisiae* cells in stationary phase from ethanol stress by reducing reactive oxygen species levels. **Yeast**, v. 33, p. 355-363, 2016.
- TEIXEIRA, M. C. *et al.* Genome-Wide Identification of *Saccharomyces cerevisiae* Genes Required for Maximal Tolerance to Ethanol. **Applied and Environmental Microbiology**, v. 75, p. 5761-5772, 2009.
- TESFAW, A., ASSEFA, F. Current trends in bioethanol production by *Saccharomyces cerevisiae*: substrate, inhibitor reduction, growth variables, coculture, and immobilization. **International Scholarly Research Notices**, 2014.
- TRAPNELL, C. *et al.* TopHat: discovering splice junctions with RNA-Seq. **Bioinformatics**, v. 25, p. 1105-1111, 2009.
- TRAPNELL, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, v. 28, p. 511-515, 2010.
- TSAI, S. Q., JOUNG, J. K. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. **Nature Reviews**, v. 17, p. 300-312, 2016.
- U.S. Department of Energy, Alternative Fuels Data Center. Global Ethanol Production, dados 2015. Disponível em <http://www.afdc.energy.gov/data/categories/biofuels-production>. Acesso 10 abr. 2017.
- VAN DER OOST, J. *et al.* Unravelling the structural and mechanistic basis of CRISPR–Cas systems. **Nature Reviews**, v. 12, p. 479-492, 2014.

- VERK, M. C. V. *et al.* RNA-Seq: revelation of the messengers. **Cell Press**, v. 18, p. 175-179, 2013.
- VOGEL, M. *et al.* Allosteric Regulation of Hsp70 Chaperones Involves a Conserved Interdomain Linker. *The Journal of Biological Chemistry*, v. 281, p. 38705-38711, 2006.
- VOORDECKERS, K. *et al.* Adaptation to High Ethanol Reveals Complex Evolutionary Pathways. **PLOS Genetics**, p. 1-31, 2015.
- WANG, Z. *et al.* RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews**, v. 10, p. 57-63, 2009.
- YAMAUCHI, Y. *et al.* Prioritized expression of *BTN2* of *Saccharomyces cerevisiae* under pronounced translation repression induced by severe ethanol stress. **Frontiers in Microbiology**, v. 7, p. 1319-1330, 2016.
- YOKOM, A. *et al.* Spiral architecture of the Hsp104 disaggregase reveals the basis for polypeptide translocation. **Nature structural and molecular biology**, v. 23, p.830-839, 2016.
- YOSHIKAWA, K. *et al.* Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in *Saccharomyces cerevisiae*. **FEMS Yeast Res**, v. 9, p. 32-44, 2009.
- YOU, K. *et al.* Ethanol tolerance in the yeast *Saccharomyces cerevisiae* is dependent on cellular oleic acid content. **Applied and Environmental Microbiology**, v. 69, p. 1499-1504, 2002.
- YUAN, T. *et al.* RNA-Seq of the xylose-fermenting yeast *Scheffersomyces stipitis* cultivated in glucose or xylose. **Applied Microbiology and Biotechnology**, v. 92, p. 1237-1249, 2011.
- ZENG, W. *et al.* Comparative transcriptomes reveal novel evolutionary strategies adopted by *Saccharomyces cerevisiae* with improved xylose utilization capability. **Applied Microbiology and Biotechnology**, v. 101, p. 1753-1767, 2017.
- ZHANG, W. *et al.* Improved ethanol production by a xylose-fermenting recombinant yeast strain constructed through a modified genome shuffling method. **Biotechnology for biofuels**, v. 5, p. 46-57, 2012.
- ZHANG, Z. *et al.* An improved method for whole protein extraction from yeast *Saccharomyces cerevisiae*. **Yeast**, v. 28, p. 795-798, 2011.
- ZHAO, X. Q. *et al.* Mechanisms of yeast stress tolerance and its manipulation for efficient fuel ethanol production. **Journal of Biotechnology**, v. 144, p. 23-30, 2009.
- ZHOU, W. *et al.* Global analyses of sumoylated proteins in *Saccharomyces cerevisiae*: Induction of protein sumoylation by cellular stresses. **Journal of Biology and Chemistry**, v. 279, p. 32262-32268, 2004.

ANEXO A

Tabela A.1- Genes expressos somente em *S. cerevisiae* Y02999

Nome do gene	FPKM Y02999_1	FPKM BY4741 média	Função*
<i>snR75</i>	5087,18	0	C/D box snoRNA ( <i>small nucleolar RNA</i> ); guia a 2'-O-metilação do rRNA presente na subunidade maior na posição G2288.
<i>RDT1</i>	3174,57	0	Fase aberta de leitura curta. Talvez seja um novo gene codificador de proteína que originou a partir de uma sequência não codificadora.
<i>snR38</i>	1560,86	0	C/D box snoRNA guia a 2'-O-metilação do rRNA presente na subunidade maior na posição G2815.
<i>YJL077W-B</i>	1025,05	0	Proteína putativa de função desconhecida.
<i>YLR264C-A</i>	292,561	0	Proteína putativa de função desconhecida.
<i>YCR095W-A</i>	39,567	0	Proteína putativa de função desconhecida.
<i>YDR098C-A</i>	15,9288	0	Retrotransposon TYA Gag co-transcrito com TYB Pol, traduzido como TYA ou poliproteína TYA-TYB. Gag é uma proteína nucleocapsídica que é o componente estrutural dos VLPs ( <i>virus-like particles</i> ), similar ao Gag retroviral.
<i>YJR128W</i>	15,2962	0	Proteína putativa de função desconhecida.
<i>YMR001C-A</i>	9,04735	0	Proteína putativa de função desconhecida.
<i>YGL015C</i>	8,33528	0	Proteína putativa de função desconhecida.
<i>YLR111W</i>	4,02176	0	Proteína putativa de função desconhecida.
<i>YHL037C</i>	2,90839	0	Proteína putativa de função desconhecida.
<i>YHL041W</i>	2,85952	0	Proteína putativa de função desconhecida.
<i>YCL075W</i>	2,11026	0	Pseudogene: codifica fragmento da proteína TY Pol.
<i>PFS1</i>	1,07872	0	Proteína de esporulação requerida para formação da membrana plasmática do esporo.

---

<i>COS2</i>	0,727951	0	Proteína endossomal envolvida no <i>turnover</i> de proteínas de membrana plasmática.
<i>YML045W</i>	51955	0	Retrotransposon de TYA Gag e genes TYB Pol, transcrito e traduzido como uma unidade, poliproteína é processada para fazer a proteína do nucleocapsídeo (Gag), transcriptase reversa, protease e integrase. Similar aos genes retrovirais.
<i>YDR210W-A</i>	0,0003299	0	Retrotransposon TYA Gag co-transcrito com TYB Pol, traduzido como TYA ou poliproteína TYA-TYB. Gag é uma proteína do nucleocapsídeo que é um constituinte estrutural dos VLPs ( <i>virus-like particles</i> ).

---

Tabela A.2- Genes expressos somente em *S. cerevisiae* BY4741

Nome do gene	FPKM Y02999_1	FPKM BY4741 média	Função
<i>FMP49</i>	0	0,00363739	Proteína mitocondrial de função desconhecida.
<i>VPS63</i>	0	0,00707415	Proteína putativa de função desconhecida.
<i>PWR1</i>	0	0,0146235	ncRNA intergênico regulatório; um dos dois ncRNAs que fazem papel central na regulação da transcrição da proteína Flo11p.
<i>YRF1-1</i>	0	0,02412925	Helicase codificada pelos Y' elementos da região subtelomérica. Altamente expressa em mutantes sem o componente da telomerase Tlc1p, potencialmente fosforilada pela Cdc28p.
<i>YER138C</i>	0	0,02929675	Retrotransposon de TYA Gag e genes TYB Pol.
<i>YLL067C</i>	0	0,03657545	Elemento putativo Y', helicase ATP dependente.
<i>COS3</i>	0	310565	Proteína endossomal envolvida no <i>turnover</i> de proteínas de membrana plasmática.
<i>YML040W</i>	0	34035	Retrotransposon TYA Gag co-transcrito com TYB Pol, traduzido como TYA ou poliproteína TYA-TYB.
<i>YHL049C</i>	0	978965	Proteína putativa de função desconhecida.
<i>YRF1-2</i>	0	9880975	Helicase codificada pelos Y' elementos da região subtelomérica. Altamente expressa em mutantes sem o componente da telomerase Tlc1p, potencialmente fosforilada pela Cdc28p.
<i>YAR010C</i>	0	0,2435455	Retrotransposon do gene TYA Gag, cotranscrito com TYB Pol.
<i>YPR137C-A</i>	0	0,2950715	Retrotransposon TYA Gag co-transcrito com TYB Pol, traduzido como TYA ou poliproteína TYA-TYB.
<i>YFL015C</i>	0	0,3999825	Proteína putativa de função desconhecida.
<i>YLL066C</i>	0	0,412928	Putativo elemento Y' helicase dependente de ATP. Não é um gene essencial.
<i>SPS4</i>	0	0,415535	Proteína cuja expressão é induzida durante a esporulação, entretanto não é requerida para esporulação. Expressão heteróloga em <i>E. coli</i> induz resposta SOS que indica dano em DNA.

<i>YFL012W</i>	0	0,471936	Proteína putativa de função desconhecida, transcrita durante esporulação.
<i>LDS1</i>	0	0,4730575	Proteína envolvida na montagem da parede de esporos, tem similaridade com Lds2p e Rrt8p.
<i>YGL118C</i>	0	0,4883875	Proteína putativa de função desconhecida. Não é um gene essencial.
<i>FDH2</i>	0	0,50794	Formato desidrogenase dependente de NAD (+). Talvez proteja as células de formato exógeno.
<i>MF(ALPHA)1</i>	0	0,510915	<i>Mating</i> feromônio de fator-alfa, feito por células alfa. Envolvido nos mecanismos que levam ao acasalamento.
<i>YFL066C</i>	0	0,516994	Proteína do tipo helicase codificada dentro de elementos teloméricos Y'.
<i>YPR064W</i>	0	0,52495	Proteína putativa de função desconhecida.
<i>YGL138C</i>	0	0,5775495	Proteína putativa de função desconhecida.
<i>RUF22</i>	0	0,623555	RNA de função desconhecida.
<i>YDR210W-C</i>	0	0,624475	Gene TYA Gag retrotransposon co-transcrito com TYB Pol. Traduzido como poliproteína TYA ou TYA-TYB.
<i>YER189W</i>	0	0,6655	Proteína putativa de função desconhecida.
<i>FDH1</i>	0	0,6728375	Formato desidrogenase dependente de NAD (+). Talvez proteja as células de formato exógeno.
<i>GAT4</i>	0	0,67609	Proteína que contém motivos <i>zinc finger</i> da família GATA, envolvida na montagem da parede do esporo.
<i>PAU11</i>	0	0,68701	Proteína putativa de função desconhecida.
<i>PAU13</i>	0	0,68701	Proteína putativa de função desconhecida. Expressão é induzida após choque por etanol.
<i>PAU20</i>	0	0,68701	Proteína putativa de função desconhecida. Expressão induzida por baixas temperaturas, em condições anaeróbicas e durante a fermentação alcoólica.
<i>YPR202W</i>	0	0,703285	Proteína putativa de função desconhecida.
<i>SNO4</i>	0	0,707245	Possível chaperona e cisteína protease. Requerida para a reprogramação transcricional durante a mudança diáuxica e para sobrevivência em fase estacionária.
<i>PAU3</i>	0	0,718065	Membro da família multigênica seripauperina. Codificado principalmente na região subtelomérica. Ativo durante a fermentação alcoólica, regulado por anaerobiose, negativamente regulado pelo oxigênio, reprimido pelo grupo heme.

<i>YEL075C</i>	0	0,72052	Proteína putativa de função desconhecida.
<i>YLR255C</i>	0	0,721935	Proteína putativa de função desconhecida.
<i>GMC2</i>	0	0,7265575	Proteína envolvida no crossing over meiótico. Componente do Complexo Sinaptonemal junto com Ecm11p. Promove a sumoilação de Ecm11p.
<i>YCR025C</i>	0	0,72895	Proteína putativa de função desconhecida.
<i>YBL044W</i>	0	0,744505	Proteína putativa de função desconhecida.
<i>YER085C</i>	0	0,8106095	Proteína putativa de função desconhecida.
<i>YDR102C</i>	0	0,81846	Proteína putativa de função desconhecida.
<i>YDR010C</i>	0	0,81846	Proteína putativa de função desconhecida.
<i>YEL009C-A</i>	0	0,856875	Proteína putativa de função desconhecida.
<i>YLR157C-B</i>	0	0,8772725	Genes retrotransposon TYA Gag e TYA Pol. Transcrito/traduzido com uma unidade. Poliproteína processada para gerar uma proteína tipo nucleocapsídeo (Gag), Transcriptase reversa (RT), Protease (PR) e integrase (IN). Similar a genes retrovirais.
<i>YAR068W</i>	0	0,896466	Proteína de função desconhecida induzida em células com deficiência respiratória.
<i>IRC13</i>	0	0,92319	Proteína putativa de função desconhecida.
<i>YPL261C</i>	0	0,963965	Proteína putativa de função desconhecida.
<i>YDR114C</i>	0	1,00828	Proteína putativa de função desconhecida. A deleção do gene leva a diminuição do crescimento em pHs elevados e elevadas concentrações de cálcio.
<i>YGR025W</i>	0	1,00828	Proteína putativa de função desconhecida.
<i>YNL143C</i>	0	1,009575	Proteína de função desconhecida.
<i>YCR022C</i>	0	1,03164	Proteína putativa de função desconhecida.
<i>YOR376W</i>	0	1,067445	Proteína putativa de função desconhecida.

<i>YJL007C</i>	0	1,0881	Proteína putativa de função desconhecida.
<i>SPO19</i>	0	1,1288935	Prosporo proteína específica de meiose, requerida para produzir a força de flexão necessária para a montagem apropriada do prosporo durante a esporulação.
<i>YAL064W</i>	0	1,16736	Proteína de função desconhecida.
<i>YOR343C</i>	0	1,18106	Proteína putativa de função desconhecida.
<i>YHL005C</i>	0	1,182365	Proteína putativa de função desconhecida.
<i>IPL1</i>	0	1,18425	Aurora quinase do complexo cromossomal passageiro. Ajuda a manter o cromossomo condensado durante a anáfase.
<i>YCR001W</i>	0	1,257445	Proteína putativa de função desconhecida.
<i>YHR125W</i>	0	1,26301	Proteína putativa de função desconhecida.
<i>SAE3</i>	0	1,26546	Proteína específica de meiose envolvida na recombinação meiótica.
<i>YDR261C-C</i>	0	1,30381	Retrotransposon TYA Gag gene cotranscrito com TYB Pol.
<i>YLR012C</i>	0	1,39992	Proteína putativa de função desconhecida.
<i>YOL085C</i>	0	1,54843	Proteína putativa de função desconhecida.
<i>HED1</i>	0	1,631035	Proteína específica da meiose, requerida para a normal morfogênese do complexo sinaptonemal.
<i>YDR157W</i>	0	1,702055	Proteína putativa de função desconhecida.
<i>YBR292C</i>	0	1,879815	Proteína putativa de função desconhecida.
<i>YGR273C</i>	0	2,044755	Proteína putativa de função desconhecida.
<i>YLR462W</i>	0	2,073975	Proteína putativa de função desconhecida similar as helicases.
<i>PAU10</i>	0	2,081195	Proteína de função desconhecida; membro da família multigênica seripauperina, codificada principalmente em regiões subtelomérica.
<i>COS12</i>	0	2,104435	Proteína endossomal envolvida no <i>turnover</i> de proteínas da membrana plasmática, membro da subfamília DUP380.
<i>YLR296W</i>	0	2,10758	Proteína putativa de função desconhecida.

<i>YIL054W</i>	0	2,215435	Não essencial proteína de função incerta. Talvez tenha uma ação como regulador negativo da transcrição de genes envolvidos no crescimento de pseudohifas.
<i>YHR139C-A</i>	0	2,21872	Proteína putativa de função desconhecida.
<i>YMR030W-A</i>	0	2,218825	Proteína putativa de função desconhecida.
<i>ECM23</i>	0	2,23836	Não essencial proteína de função incerta. Talvez tenha uma ação como regulador negativo da transcrição de genes envolvidos no crescimento de pseudohifas.
<i>YFL064C</i>	0	2,440855	Proteína putativa de função desconhecida.
<i>YKL102C</i>	0	2,46969	Proteína putativa de função desconhecida.
<i>HES1</i>	0	2,476655	Proteína envolvida na regulação da biossíntese de ergosterol.
<i>YMR082C</i>	0	2,54756	Proteína putativa de função desconhecida.
<i>YLR112W</i>	0	2,701275	Proteína putativa de função desconhecida.
<i>AHT1</i>	0	2,705325	Proteína putativa de função desconhecida. Supressor multicópia de defeitos de transporte de glicose, provavelmente devido a presença de elemento regulatório Hxt4p.
<i>CSM4</i>	0	2,73138	Proteína requerida para segregação acurada dos cromossomos durante a meiose.
<i>YER084W</i>	0	2,7664	Proteína de função desconhecida.
<i>YOR008C-A</i>	0	2,826385	Proteína putativa de função desconhecida; incluí um domínio potencial de transmembrana. Deleção resulta em telômero levemente aumentados.
<i>YPL080C</i>	0	2,883105	Proteína putativa de função desconhecida.
<i>YKR073C</i>	0	2,99982	Proteína putativa de função desconhecida.
<i>VEL1</i>	0	3,094555	Proteína de função desconhecida. Fortemente induzida em condições de depleção de zinco e tem expressão aumentada em mutantes com <i>NAPI</i> deletado.
<i>YHR130C</i>	0	3,273295	Proteína putativa de função desconhecida, conservada em linhagens de <i>S. cerevisiae</i> . Não é um gene essencial.

<i>API2</i>	0	3,354885	Proteína putativa de função desconhecida, conservada em linhagens de <i>S. cerevisiae</i> .
<i>YDL022C-A</i>	0	3,363595	Proteína de função desconhecida.
<i>YDR246W-A</i>	0	3,398975	Proteína putativa de função desconhecida.
<i>YGL235W</i>	0	3,56509	Proteína putativa de função desconhecida, potencial substrato de Cdc28p. Mutantes nulos apresentam aumento de resistência aos agentes antifúngicos gliotoxina, cicloheximida e H <sub>2</sub> O <sub>2</sub> .
<i>JIP3</i>	0	3,684745	Proteína putativa de função desconhecida, conservada entre linhagens de <i>S. cerevisiae</i> .
<i>YLR236C</i>	0	3,846595	Proteína putativa de função desconhecida.
<i>YDL177C</i>	0	4,14946	Proteína putativa de função desconhecida.
<i>YPL278C</i>	0	4,231745	Proteína putativa de função desconhecida.
<i>YMR254C</i>	0	4,307325	Proteína putativa de função desconhecida.
<i>YER039C-A</i>	0	5,00125	Proteína putativa de função desconhecida.
<i>YAL063C-A</i>	0	5,08117	Proteína putativa de função desconhecida.
<i>YOR072W</i>	0	5,138775	Proteína putativa de função desconhecida.
<i>YBL039W-A</i>	0	5,33065	Proteína putativa de função desconhecida.
<i>YLR053C</i>	0	5,709495	Proteína putativa de função desconhecida.
<i>PAU5</i>	0	5,831985	Membro da família multigênica das seripauperinas. Codificado principalmente em regiões subteloméricas, ativa durante fermentação alcoólica, regulado por anaerobiose, negativamente regulado por oxigênio, reprimida por heme.
<i>YCR041W</i>	0	5,914795	Proteína de função desconhecida. Sua superexpressão suprime as perdas de alta frequência de mini-cromossomos, provavelmente por aumentar a taxa de segregação adequada de cromossomos.
<i>LSO1</i>	0	5,9923	Proteína com papel potencial na resposta a privação de ferro. Transcrição aumenta durante privação de ferro e durante tratamento com cobre.

<i>PAU17</i>	0	6,26609	Proteína de função desconhecida, membro da família multigênica das seripauperinas. Codificado principalmente em regiões subteloméricas. Não é um gene essencial.
<i>YFL067W</i>	0	6,476675	Proteína de função desconhecida; <i>down</i> -regulada em baixas concentrações de cálcio.
<i>YBR032W</i>	0	6,71019	Proteína putativa de função desconhecida.
<i>YIR020C</i>	0	7,21603	Proteína de função desconhecida.
<i>YGR240C-A</i>	0	7,287115	Proteína putativa de função desconhecida.
<i>YFL041W-A</i>	0	8,3932	Proteína putativa de função desconhecida.
<i>YCR050C</i>	0	8,85901	Proteína não-essencial de função desconhecida.
<i>LOT6</i>	0	9,228555	FMN-dependente NAD(P)H: quinona redutase, desempenha papel em morte celular do tipo apoptose. Provavelmente envolvida em desintoxicação de quinonas. Expressão elevada em baixas temperaturas. Em condições redutoras, sequestra o fator de transcrição Cin5p no citoplasma em complexo com o proteassomo.
<i>YOL097W-A</i>	0	9,2613	Proteína putativa de função desconhecida.
<i>YBR200W-A</i>	0	10,2944	Proteína putativa de função desconhecida.
<i>YOL159C-A</i>	0	11,52025	Proteína de função desconhecida, superexpressão afeta o tráfico de proteínas endocíticas.
<i>YPR195C</i>	0	12,038475	Proteína putativa de função desconhecida.
<i>YOR011W-A</i>	0	12,173205	Proteína putativa de função desconhecida.
<i>YBR072C-A</i>	0	12,28395	Proteína putativa de função desconhecida.
<i>YHR175W-A</i>	0	13,11325	Proteína putativa de função desconhecida.
<i>YOR034C-A</i>	0	13,1395	Proteína putativa de função desconhecida.
<i>YOR105W</i>	0	14,32743	Proteína de função desconhecida.
<i>YBR230W-A</i>	0	14,49495	Proteína putativa de função desconhecida.
<i>YOR161C-C</i>	0	14,68235	Proteína de função desconhecida.

---

<i>YHR007C-A</i>	0	17,97648	Proteína putativa de função desconhecida.
<i>APJ1</i>	0	18,7887	Chaperona com papel na degradação de proteínas sumoiladas. Membro da família do tipo DnaJ.
<i>YDR034C-A</i>	0	21,17265	Proteína putativa de função desconhecida.
<i>MHF2</i>	0	22,0377	Componente do complexo heterotetrâmico MHF envolvido no dobramento de histonas.
<i>YKL065W-A</i>	0	22,34615	Proteína putativa de função desconhecida.
<i>PAI3</i>	0	27,98305	Inibidor da proteinase citoplasmática A (Pep4p).
<i>YJL047C-A</i>	0	46,0196	Proteína putativa de função desconhecida.
<i>YOR376W-A</i>	0	49,15125	Proteína putativa de função desconhecida.
<i>YBL101W-C</i>	0	124,328	Proteína putativa de função desconhecida.
<i>RDN5-1</i>	0	140,9415	RNA ribossomal 5S, componente da subunidade ribossomal 60S. Transcrição é mediada pela PolIII e ativada por TFIIIA e TFIIIE. Talvez tem um papel na fidelidade da tradução.
<i>snR56</i>	0	656,56	C/D box snoRNA ( <i>small nucleolar RNA</i> ). Guia a 2'-O –metilação presente na subunidade menor rRNA na posição G1428.
<i>YOR302W</i>	0	1977,635	Peptídeo atenuador de arginina. Regula a tradução do mRNA Cpa1p.
<i>snR24</i>	0	9155,245	C/D box snoRNA ( <i>small nucleolar RNA</i> ). Guia a metilação da subunidade maior rRNA nas posições C1437, C1449 e C1450.

---

Tabela A.3- Genes expressos em *S. cerevisiae* BY4741 e Y02999

Nome do gene	FPKM Y02999_1	FPKM BY4741 média	FC (Y02999_1/ BY4741 média)	log2 FC	Função
<b>Genes <i>up</i>-regulados</b>					
<i>IME4</i>	6,37416	0,00138841	4590,9947	12,1646	mRNA metiltransferase N6-adenosina, requerida para entrada em meiose, media a metilação de N6-adenosina nos reguladores de meiose IME1, IME2 e IME4 durante a indução da esporulação.
<i>NAG1</i>	2,615	0,0026627	982,0877	9,9397	Proteína envolvida na biogênese da parede celular de leveduras.
<i>YBL101W-A</i>	0,82883	0,051502	16,0932	4,0084	Retrotransposon TYA Gag co-transcrito com TYB Pol.
<i>YDR278C</i>	8,05195	0,904005	8,9070	3,1549	Proteína putativa de função desconhecida.
<i>JJJ3</i>	11,7777	1,486505	7,9231	2,9861	Proteína de função desconhecida. Contém um domínio DnaJ e um domínio CSL dedo de zinco. Envolvida na biossíntese de diftamida.
<i>PRM6</i>	10,8021	1,48295	7,2842	2,8648	Transportador de potássio que medeia o influxo de K <sup>+</sup> . Ativa o sistema de influxo de Ca <sup>2+</sup> durante resposta ao feromônio de conjugação. Expressão <i>up</i> -regulada em resposta ao fator alfa.
<i>YNL042W-B</i>	14,0921	1,96339	7,1774	2,8435	Proteína putativa de função desconhecida.
<i>YOR032W-A</i>	27,6781	4,134905	6,6938	2,7428	Proteína de função desconhecida.
<i>NAT4</i>	2,21251	0,3712485	5,9596	2,5752	N-alfa-acetil-transferase, envolvida na acetilação dos resíduos N-terminais das histonas H4 e H2A.
<i>RPM1</i>	67,7422	11,82268	5,7299	2,5185	RNA componente da RNase P mitocondrial.

<i>YGR050C</i>	20,8207	3,97983	5,2316	2,3872	Proteína de função desconhecida.
<i>YIR018C-A</i>	109,082	21,27865	5,1264	2,3579	Proteína putativa de função desconhecida.
<i>Q0255</i>	9,20298	1,85715	4,9554	2,3090	Proteína do tipo maturase.
<i>YGL149W</i>	9,75855	1,97131	4,9503	2,3075	Proteína putativa de função desconhecida.
<i>YLR307C-A</i>	7,03012	1,42208	4,9435	2,3055	Proteína putativa de função desconhecida.
<i>YDL241W</i>	9,68886	1,965675	4,9290	2,3013	Proteína putativa de função desconhecida.
<i>YCL001W-B</i>	7,56692	1,564275	4,8373	2,2742	Proteína putativa de função desconhecida.
<i>YMR175W-A</i>	18,3028	3,82087	4,7902	2,2601	Proteína putativa de função desconhecida.
<i>YNL162W-A</i>	11,4996	2,500615	4,5987	2,2012	Proteína putativa de função desconhecida.
<i>AIR1</i>	8,11985	1,766405	4,5968	2,2006	Proteína <i>zinc knuckle</i> . Envolvida no processamento e degradação de RNA nuclear, por ser componente do complexo TRAMP.
<i>YFR032C-B</i>	6,52371	1,42208	4,5874	2,1977	Proteína putativa de função desconhecida.
<i>SRL4</i>	2,18157	0,4801	4,5440	2,1840	Proteína de função desconhecida, envolvida na regulação da produção de dNTP.
<i>YHL009W-A</i>	1,32135	0,29556865	4,4705	2,1604	Retrotransposon TYA Gag co-transcrito com TYB Pol, transcrito como TYA ou poliproteína TYA-TYB.
<i>JLP2</i>	10,3981	2,39324	4,3448	2,1193	Proteína de função desconhecida, contem sequência que se assemelha a um domínio J.
<i>YJL218W</i>	5,56308	1,3034985	4,2678	2,0935	Putativa acetiltransferase.
<i>FYV12</i>	2,51943	0,59957	4,2021	2,0711	Proteína de função desconhecida, requerida para sobrevivência quando há exposição à toxina K1.
<i>tS(UGA)Q2</i>	2206,63	538,015	4,1014	2,0361	Serina mitocondrial tRNA ( <i>tRNA-Ser</i> ).

<i>ABM1</i>	5,70218	1,42775	3,9938	1,9978	Proteína de função desconhecida, requerida para organização adequada dos microtúbulos.
<i>SLD2</i>	4,03791	1,013957	3,9823	1,9936	Proteína de anelamento e ligação a origem de replicação. Requerida para início da replicação de DNA. Fosforilada na fase S por ciclinas dependentes de quinases (cdks). Envolvida na formação do complexo Dpb11. Componente do complexo de pré-carregamento. Se liga ao complexo Mcm2-7 para prevenir interação inapropriada com o complexo GINS durante a fase G1.
<i>YML133C</i>	0,126668	0,031874	3,9740	1,9906	Putativo Y' elemento helicase ATP-dependente.
<i>PAU12</i>	5,79436	1,519975	3,8121	1,9306	Proteína de função desconhecida, membro da família multigênica das seripauperinas, codificado principalmente em regiões subteloméricas.
<i>ECM11</i>	0,828157	0,225053	3,6798	1,8796	Proteína meiose-específica, componente do complexo sinaptonemal junto com Gmc2p, requerida para formação eficiente do crossover e para o carregamento eficiente da proteína Zip1p. É sumoilada de uma maneira dependente de Gmc2p e a sumoilção é essencial para sua função na meiose.
<i>YKR041W</i>	2,69828	0,734992	3,6712	1,8762	Proteína de função desconhecida.
<i>TOM5</i>	388,855	106,7796	3,6417	1,8646	Componente do complexo TOM (translocase da membrana exterior), responsável pelo reconhecimento e importação inicial de todas as proteínas direcionadas para as mitocôndrias.
<i>CWC15</i>	10,4811	2,88553	3,6323	1,8609	Proteína não-essencial envolvida no <i>splicing</i> de pré-mRNA.

<i>OST4</i>	342,741	94,5205	3,6261	1,8584	Subunidade do complexo oligossacariltransferase do lúmen do retículo endoplasmático. Este complexo catalisa a glicosilação de proteínas ligadas a asparagina. Proteína de membrana Tipo I requerida para incorporação de Ost3p ou Ost6p dentro do complexo OST.
<i>MF(ALPHA)2</i>	13,3	3,72408	3,5714	1,8365	Feromônio de conjugação do fator alfa, feito pelas células alfa.
<i>YIL086C</i>	9,38674	2,64168	3,5533	1,8292	Proteína putativa de função desconhecida.
<i>YKR032W</i>	4,03209	1,14092	3,5341	1,8213	Proteína putativa de função desconhecida.
<i>YBR144C</i>	4,54979	1,3044	3,4880	1,8024	Proteína putativa de função desconhecida.
<i>YDL114W</i>	2,51544	0,727027	3,4599	1,7907	Desidrogenase/redutase putativa de cadeia curta, não é um gene essencial.
<i>YMR244W</i>	4,90174	1,435585	3,4145	1,7717	Proteína putativa de função desconhecida.
<i>SPG3</i>	33,6503	9,87989	3,4059	1,7681	Proteína requerida para sobrevivência a altas temperaturas durante a fase estacionária. Não é necessária para o crescimento em fontes de carbono não-fermentáveis.
<i>SRT1</i>	3,92705	1,1552365	3,3993	1,7653	Forma o complexo diidrodicol difosfato sintase (DDS) com Nus1p. Envolvida na biossíntese de dolícois de cadeias longas. Transcrição é induzida na fase estacionária.
<i>DCVI</i>	4,84052	1,43251	3,3790	1,7566	Proteína de função desconhecida.
<i>VARI</i>	15,5223	4,73972	3,2749	1,7115	Proteína ribossomal mitocondrial da subunidade menor. Talvez desempenha um papel na perda de DNA mitocondrial em condições de estresse.

<i>IMP3</i>	21,9204	6,702895	3,2703	1,7094	Componente do SSU processomo, necessário para o processamento do pré-18S rRNA. Interage com Mpp10p e medeia a interação de Iml4p e Mpp10p com U3 snoRNA.
<i>SPR28</i>	1,10426	0,345612	3,1951	1,6759	Proteína específica de esporulação, homóloga da família de genes CDC3/10/11/12. Expressa em altos níveis durante a divisão meiótica e na formação de ascósporos.
<i>YFL019C</i>	14,9058	4,673795	3,1892	1,6732	Proteína putativa de função desconhecida.
<i>SPS1</i>	7,31678	2,37395	3,0821	1,6239	Putativa serina/treonina quinase. Necessária para o eficiente empacotamento de esporos, desenvolvimento da membrana do prosporo e localização de enzimas envolvidas na síntese da parede do esporo. É requerida para a fosforilação e <i>turnover</i> de Ssp1. Expressa no final da meiose.
<i>YAP6</i>	10,5959	3,45489	3,0669	1,6168	Fator de transcrição bZIP ( <i>basic leucine zipper</i> ). Interage fisicamente com complexo Tup1-Cyc8 e recruta Tup1 para seus alvos. Análises computacionais sugerem um papel na regulação da expressão de genes envolvidos no metabolismo de carboidratos. Ativador transcricional envolvida na regulação de genes expressos em resposta a mudanças ambientais e necessidades metabólicas. Regula genes envolvidos na biogênese de ribossomos, síntese de proteínas, metabolismo e transporte de carboidratos. Sua superexpressão confere resistência ao metilmetanosulfato e mitomicina e aumenta a tolerância a sódio e lítio.

<i>RME3</i>	5,47882	1,79465056	3,0529	1,6102	Transcrito anti-senso que reprime o gene <i>ZIP2</i> meiose-específico. Expresso em condições que induzem a esporulação.
<i>ARG3</i>	8,5785	2,82892	3,0324	1,6005	Ornitina carbamoiltransferase. Catalisa a biossíntese de citrulina, precursor da arginina.
<i>ECM12</i>	13,7358	4,576415	3,0014	1,5857	Proteína putativa de função desconhecida, talvez contribua para a biossíntese da parede celular.
<i>YBR056W-A</i>	140,202	47,7114	2,9385	1,5551	Proteína de função desconhecida.
<i>BTN2</i>	27,3663	9,3157	2,9377	1,5547	Proteína de ligação v-SNARE. Modula a captação de arginina. Tem possível papel na mediação da homeostase de pH entre o vacúolo e a membrana plasmática. Expresso preferencialmente em casos de severo estresse etanólico.
<i>YGL262W</i>	4,43407	1,50987	2,9367	1,5542	Proteína putativa de função desconhecida, não é um gene essencial.
<i>YEL008W</i>	7,32617	2,517295	2,9103	1,5412	Proteína putativa de função desconhecida, conservada entre linhagens de <i>S. cerevisiae</i> . Não é um gene essencial. Envolvido em metabolismo.
<i>YDR344C</i>	7,57317	2,635315	2,8737	1,5229	Proteína putativa de função desconhecida.
<i>YJR029W</i>	0,272068	0,0952335	2,8569	1,5144	Retrotransposon TYA Gag e TYB Pol.
<i>RPA12</i>	25,0025	8,774555	2,8494	1,5107	Subunidade A12.2 da RNA polimerase I, contém dois domínios de ligação a zinco e um domínio N terminal responsável pela ancoragem do complexo da RNA pol I. Interage fisicamente com o ativador transcricional Msn4 para regular a transcrição de <i>AYRI</i> , um gene envolvido no metabolismo de lipídeo.

---

<b>Genes <i>down</i>-regulados</b>					
<i>ULI1</i>	3,76877	10,90125	0,3457	-1,5323	Proteína de função desconhecida, induzida pela resposta de proteínas não dobradas do retículo endoplasmático (UPR).
<i>HXT12</i>	4,14379	12,2318	0,3388	-1,5616	Possível pseudogene. Codifica uma proteína não funcional membro da família de transportadores de hexoses.
<i>ATP23</i>	1,83809	5,453485	0,3370	-1,5690	Putativa metaloprotease da membrana interna da mitocôndria. Necessária para o processamento de Atp6 e tem um papel na montagem do setor FO do complexo F1FO ATP sintase.
<i>RPR2</i>	2,17328	6,468585	0,3360	-1,5736	Subunidade da RNase P nuclear. RNase P cliva o precursor de tRNA para gerar a extremidade 5' madura e facilita o <i>turnover</i> de RNAs nucleares. A abundância da proteína aumenta em resposta a estresses durante a replicação do DNA.
<i>YBR013C</i>	3,15705	9,434395	0,3346	-1,5794	Proteína putativa de função desconhecida.
<i>SMA1</i>	1,23014	3,708405	0,3317	-1,5920	Proteína de função desconhecida envolvida da montagem da membrana do prosporo durante a esporulação. Interage com Spo14p.
<i>TSC3</i>	26,3928	79,6507	0,3314	-1,5935	Proteína que estimula a atividade da serina palmitoiltransferase. Envolvida na biossíntese de esfingolipídios.
<i>HSK3</i>	12,4022	37,6351	0,3295	-1,6015	Subunidade do complexo dam1. Auxilia na segregação do cromossomo durante a mitose.

---

<i>UNG1</i>	2,75279	8,390135	0,3281	-1,6078	Uracil-DNA glicosilase. Requerida para reparar a presença de uracila no DNA formada pela desaminação espontânea da citosina. Não é necessária para correção de mal pareamentos.
<i>YBR138C</i>	1,34256	4,165225	0,3223	-1,6334	Proteína citoplasmática de função desconhecida. Potencialmente fosforilada por Cdc28p. Substrato para Cdh1p.
<i>YLR122C</i>	3,07547	9,94832	0,3091	-1,6936	Proteína putativa de função desconhecida.
<i>ECO1</i>	0,779842	2,54244	0,3067	-1,7050	Acetiltransferase. Requerida para estabilização da coesão das cromátides irmãs. Acetila Mps3p para regular a organização nuclear, modifica Smc3 na forquilha de replicação e modifica Mcd1p em resposta a quebras de dupla fita no DNA. A distribuição no núcleo aumenta em condições de estresse na replicação do DNA.
<i>NNF1</i>	2,23396	7,495445	0,2980	-1,7464	Componente essencial do complexo MIND no cinetócoro. Requerido para segregação do cromossomo.
<i>SPO13</i>	0,914738	3,120895	0,2931	-1,7705	Regulador da meiose. Envolvido na manutenção da coesão entre as cromátides irmãs durante a meiose I, assim como promove a ligação dos cinetócoros no fuso durante a meiose I e meiose II.
<i>TOS4</i>	2,40606	8,240785	0,2920	-1,7761	Putativo fator de transcrição. Envolvido na homeostase da expressão, equilibrando a taxa de síntese de mRNA com as mudanças de dosagem gênica durante a fase S. Envolvido nas respostas de checkpoints da replicação do DNA. A distribuição no núcleo aumenta em condições de estresse na replicação do DNA.

<i>IMG2</i>	3,97778	13,7123	0,2901	-1,7854	Proteína ribossomal mitocondrial, presente na subunidade maior do ribossomo.
<i>YJL225C</i>	0,101313	0,350723	0,2889	-1,7915	Putativo Y' elemento helicase ATP-dependente.
<i>SOH1</i>	8,28401	28,7967	0,2877	-1,7975	Subunidade do complexo mediador da RNA polimearse II. Associa com o core da polimerase para formar a holoenzima. Envolvida na manutenção do telômero.
<i>YDR261C-D</i>	1,41573	5,028625	0,2815	-1,8286	Retrotransposon TYA Gag e TYB Pol genes.
<i>YHR214W</i>	1,266	4,57516	0,2767	-1,8535	Proteína putativa de função desconhecida. Predita para ser uma proteína modificada de glicosilfosfatidilinositol.
<i>RRT1</i>	5,06274	18,4091	0,2750	-1,8624	Proteína putativa de função desconhecida.
<i>AIM11</i>	4,71281	17,2207	0,2737	-1,8695	Proteína de função desconhecida. Mutantes nulos deste gene apresentam aumento na perda do genoma mitocondrial.
<i>MET28</i>	2,88152	10,68345	0,2697	-1,8905	Ativador transcricional bZIP no complexo Cbf1-Met4-Met28. Participa da regulação do metabolismo de enxofre.
<i>FMP32</i>	2,53707	9,478425	0,2677	-1,9015	Fator de montagem putativo para citocromo <i>c</i> oxidase.
<i>PEX17</i>	1,49601	6,12422	0,2443	-2,0334	Peroxina de membrana do complexo importador peroxissomal. Este complexo facilita a importação de proteínas necessárias para biogênese do peroxissomo.
<i>IRC18</i>	1,19881	5,003045	0,2396	-2,0612	Proteína envolvida na montagem da parede externa do esporo. Expressão é induzida em células com problemas de respiração e em culturas de quimiostato com carbono limitante.

<i>MIM2</i>	4,73166	20,896265	0,2264	-2,1428	Proteína mitocondrial necessária para importação de proteínas da membrana externa. Importação de proteínas com múltiplos segmentos alfa-hélices transmembrana, incluindo Ugo1p, Tom20p e Fzo1p.
<i>SEN34</i>	1,74305	7,842045	0,2223	-2,1696	Subunidade da endonuclease tRNA <i>splicing</i> que foram o complexo Sen responsável por clivar o mRNA CBP1 na superfície da mitocôndria.
<i>YNL092W</i>	1,11611	5,38586	0,2072	-2,2707	Proteína metiltransferase dependente de S-adenosilmetionina. Capacidade de autometilação.
<i>BUD31</i>	2,12319	10,478505	0,2026	-2,3031	Componente do subcomplexo SF3b da U2 snRNP. Aumenta a eficiência do primeiro e segundo passo do <i>splicing</i> . Facilita a passagem do início da fase G1/S mas não é necessária para G2/M ou saída da mitose.
<i>CSE4</i>	2,05082	10,12238	0,2026	-2,3033	Proteína centromérica similar a histona H3. Necessária para função apropriada do cinetócoro. Ubiquitinação da extremidade N-terminal regula a proteólise para a fiel segregação dos cromossomos. Níveis regulados pela E3 ubiquitina ligase Psh1p.
<i>SMX2</i>	8,3274	44,0516	0,1890	-2,4033	Proteína do core Sm, parte do complexo heterohepatamerico que é parte dos snRNPs spliceossomal (U1, U2, U4 e U5).
<i>COX23</i>	3,71209	20,4747	0,1813	-2,4635	Proteína que mantém a homeostase de cobre na mitocôndria. Proteína do espaço mitocondrial intermembrana. Essencial para expressão de citocromo oxidase.
<i>DIG2</i>	1,4007	8,330375	0,1681	-2,5722	Inibidor responsivo a MAP-quinase do fator de transcrição Ste12p. Envolvido na regulação de genes específicos do acasalamento.
<i>BNS1</i>	2,3262	13,916745	0,1672	-2,5808	Proteína de função desconhecida.

<i>DIT2</i>	0,498234	3,630495	0,1372	-2,8653	N-formiltirosina oxidase; enzima microsomal envolvida na produção de N,N- bisformil ditirosina requerida para maturação da parede do esporo.
<i>BBP1</i>	0,507508	4,70257	0,1079	-3,2119	Proteína requerida para duplicação do centrôssomo ( <i>spindle pole body</i> ). Forma um complexo com a proteína de envelope nuclear Msp2p e com Spc29p e Kar1p. Necessária para função mitótica de Cdc5p.
<i>YDR316W-B</i>	1,18023	10,999332	0,1073	-3,2203	Retrotransposon TYA Gag e TYB Pol genes.
<i>POA1</i>	1,58848	15,33835	0,1036	-3,2714	Fosfatase altamente específica para ADP-ribose-1-fosfatase. Talvez tenha um papel na regulação do <i>splicing</i> de tRNA.
<i>YBR012W-A</i>	0,49493	11,33169	0,0437	-4,5170	Retrotransposon TYA Gag gene co-transcrito com TYB Pol.
<i>VAM10</i>	0,0044788	0,3435225	0,0130	-6,2611	Proteína envolvida na morfogênese do vacúolo.

\*Funções retiradas do *Saccharomyces Genome Database*