
Ciências Biológicas

André Young Ferreira

**Estudo de expressão diferencial do
transcriptoma de *Saccharomyces cerevisiae*
(Linhagem Pedra 2) em diferentes
concentrações de etanol**

André Young Ferreira

Estudo de expressão diferencial do transcriptoma de *Saccharomyces cerevisiae*
(Linhagem Pedra 2) em diferentes concentrações de etanol

Orientador: Dr. Jeferson Gross

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho” - Câmpus de Rio Claro, para obtenção do grau de bacharel em Ciências Biológicas.

Rio Claro
2018

F383e Ferreira, André Young
 Estudo de expressão diferencial do transcriptoma de
Saccharomyces cerevisiae (Linhagem Pedra 2) em
diferentes concentrações de etanol / André Young
Ferreira. -- Rio Claro, 2018
 25 p. : il., tabs.

 Trabalho de conclusão de curso (Bacharelado - Ciências
Biológicas) - Universidade Estadual Paulista (Unesp),
Instituto de Biociências, Rio Claro
 Orientador: Jeferson Gross

 1. Saccharomyces cerevisiae. 2. Transcriptoma. 3.

Resistência ao Etanol. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do
Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Resumo

Leveduras da espécie *Saccharomyces cerevisiae* são amplamente utilizadas na indústria da produção de etanol pela fermentação alcoólica da cana-de-açúcar (*Saccharum sp*). A fermentação é realizada em biorreatores, e uma estratégia bastante difundida para se economizar dinheiro e tempo é reutilizar as células como inóculo por vários ciclos de fermentação (reciclo). Linhagens selvagens de *S. cerevisiae*, isto é, que naturalmente habitam a cana-de-açúcar, tem a capacidade de sobreviver a esses ciclos fermentativos, podendo ser utilizadas em reciclo. Uma das linhagens selvagens mais utilizadas pela indústria brasileira é a Pedra 2 (PE-2), linhagem estudada neste trabalho, em que foi feita uma análise transcriptômica de condições normais e em estresse de etanol (6% v/v etanol). O conjunto dos RNAs de uma célula é denominado transcriptoma. Para obtê-lo se utilizam métodos de sequenciamento de RNA. Sabendo a quantidade de diferentes sequências de mRNAs em uma amostra, é possível se ter informações sobre a expressão gênica da amostra no momento da coleta. Com esses dados, análises estatísticas foram utilizadas para comparar diferentes amostras, e compreender as relações e funcionalidades dos genes diferencialmente expressos. A resposta de *S. cerevisiae* ao estresse de álcool é bastante complexa, e dificilmente se encontra um padrão de expressão gênica durante o estresse. Os resultados obtidos, após filtragem dos dados por significância, demonstraram regulação positiva de 214 genes, e negativa de 355 genes. Das funções analisadas, genes relacionados à ribossomos foram os mais afetados negativamente, porém genes ligados à biossíntese de ribossomos tiveram uma expressividade positiva. A regulação negativa de genes relacionados a ribossomos e proteínas demonstra resposta ao estresse de etanol, porém a resposta obtida quanto aos mecanismos de tolerância ao etanol por parte de genes relacionados a membrana plasmática, parede celular, proteínas de choque térmico, regeneração de NADH/NADPH, gliconeogênese, e alguns fatores de transcrição não demonstraram correlação com o que já havia sido descrito na literatura em outras linhagens. Isso pode ser devido a um diferente mecanismo de resposta ao etanol por PE-2, ou simplesmente por falta de estudos suficientes sobre a resposta específica ao etanol por *S. cerevisiae*.

SUMÁRIO

1. INTRODUÇÃO.....	4
1.1. Rna-Seq e análise de expressão gênica diferencial.....	4
1.2. Análise de representatividade funcional de genes.....	5
1.3. Tolerância ao etanol em <i>S. cerevisiae</i>	6
2. OBJETIVOS.....	8
3. MATERIAIS E MÉTODOS.....	8
3.1. Cultivo e obtenção do transcriptoma de <i>S. cerevisiae</i> PE-2.....	8
3.2. Filtragem e mapeamento de sequencias contra modelos gênicos.....	8
3.3. Estudo de expressão gênica diferencial.....	9
3.4. Análise GO.....	10
4. RESULTADOS.....	10
4.1. Estudo de expressão gênica diferencial.....	10
4.2. Análise GO.....	13
5. DISCUSSÃO.....	15
6. CONCLUSÃO.....	19
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	20

1. Introdução

1.1 RNA-Seq e análise de expressão gênica diferencial

Transcriptoma se trata do conjunto dos RNAs de uma, ou de um grupo de células, transcritos pelas enzimas RNA Polimerases (RNAPol) a partir de informações contidas no DNA. Deste modo, o transcriptoma engloba de forma quantitativa e qualitativa todo o RNA encontrado encontrados em uma célula ou em um agrupamento celular em um momento ou condição específica. (BROWN, 2002) A composição destes transcritos inclui RNAs mensageiros (mRNA), RNAs não codificantes e moléculas de *small* RNA (sRNA). Estimando-se a quantidade de cada transcrito relativa ao total de transcritos, é possível comparar mudanças na expressão de genes de amostras obtidas em diferentes condições ou estágios de desenvolvimento (WANG et al., 2009).

Existem diversos métodos para se obter um transcriptoma, porém os mais utilizados atualmente se baseiam em sequenciamentos de alto rendimento, ou seja, o que é conhecido como High-Throughput Sequencing (HTS), aplicado frequentemente no sequenciamento de moléculas de DNA (WANG et al., 2009). O método RNA-Seq, baseado em HTS, possui algumas vantagens relevantes em comparação a métodos baseados em hibridização, como Microarray. Métodos que utilizam hibridização para obter o transcriptoma necessitam de um conhecimento prévio do genoma para serem aplicados. Outro problema é a geração espontânea de híbridos cruzados, criando uma maior dificuldade de detecção das bases (OKONIEWSKI & MILLER. 2006; ROYCE et al., 2007). A tecnologia usada em RNA-Seq não possui uma limitação quanto a detecção apenas de sequências genômicas já conhecidas, o que faz com que RNA-Seq possa ser utilizado para o estudo de organismos não-modelo. Pela natureza fragmentada da leitura, sequências curtas podem revelar informações sobre a conectividade de exons, e também por isso é raro que ocorra hibridização cruzada. Além disso variações das sequências podem ser encontradas (MARIONI et al., 2008).

O procedimento aplicado para o sequenciamento de RNA em equipamentos de HTS geralmente se baseia na obtenção prévia de fragmentos de DNA complementar (cDNA) da amostra, isto é, moléculas de DNA obtidas a partir da transcrição reversa de um mRNA pela enzima transcriptase reversa (SONESON & DELORENZI, 2013). Após o sequenciamento, as sequências dos fragmentos são armazenadas em um computador, indistintamente e juntas em um mesmo arquivo. A partir disso, para se estimar a expressão gênica diferencial entre

amostras, é necessário que as sequências sejam alinhadas contra um genoma de referência. Isso é feito pela verificação da similaridade entre cada sequência de fragmento e o genoma de referência, geralmente já contendo a anotação dos genes, ou seja, a localização de tais genes em meio ao genoma completo e seus cromossomos. A esse processo é dado o nome de mapeamento, e seu resultado é uma contagem numérica da quantidade de fragmentos que são relacionados a cada gene da referência.

Após o sequenciamento e mapeamento, se procuram genes diferencialmente expressos, ou seja, genes que apresentem uma quantificação de expressão discordante entre os tratamentos analisados. Para comparar as amostras e encontrar genes diferencialmente expressos é necessário utilizar análises estatísticas sobre os dados obtidos pelo mapeamento, porém realizar uma análise estatística diferencial dos dados obtidos a partir de um RNA-Seq possui algumas dificuldades. (SONESON & DELORENZI, 2013) Um exemplo é o problema de que genes de maior tamanho (com mais bases) criam uma maior probabilidade de que um fragmento seja contabilizado mais vezes durante o mapeamento, se comparado com o mapeamento de um fragmento de mesma expressividade, mas contra um gene menor (OSHLACK et al., 2010). Outro problema é o fato de diferentes amostras terem diferentes quantidades de leituras (*reads*) mapeadas, ou seja, as amostras não são diretamente comparáveis, criando a necessidade de se realizar uma normalização dos dados de todas as amostras antes de prosseguir com a análise (BULLARD et al., 2010). Por conta da grande quantidade de variáveis a serem levadas em consideração na hora de gerar esse tipo de análise estatística, e das dificuldades encontradas, diferentes métodos vem sendo desenvolvidos. Um dos conjuntos de testes estatísticos mais utilizados atualmente é o DESeq2 (LOVE et al., 2014), disponível gratuitamente dentro do pacote Bioconductor para a plataforma R.

1.2 Análise de representatividade funcional de genes

“*Gene Ontology*” (GO) é uma iniciativa criada pelo *Gene Ontology Consortium* como meio de criar um vocabulário padronizado que represente genes e produtos gênicos de todas as espécies, com suas respectivas informações. (THE GENE ONTOLOGY CONSORTIUM, 2000). Esse vocabulário se expande em três categorias: (1) processo biológico, (2) função molecular e (3) componente celular. A primeira categoria, processo biológico, se refere à contribuição que certo gene realiza a um objetivo biológico, geralmente transformações físicas ou químicas de moléculas. Função molecular se refere à atividade

bioquímica de um produto gênico. Componente celular se trata do local físico dentro de uma célula onde o produto desse gene é ativo. Utilizando apenas o nome padronizado de genes de interesse é possível obter sua classificação em qualquer uma das três categorias de GO, e assim entender a relação entre diversos genes simultaneamente.

1.3 Tolerância ao etanol em *Saccharomyces cerevisiae*

Leveduras, principalmente *Saccharomyces cerevisiae*, são amplamente utilizadas em processos de fermentação, e consequentemente na produção de etanol a partir de açúcares extraídos de vegetais, como a cana-de-açúcar (*Saccharum sp*). Uma estratégia bastante usada durante a produção de etanol é, ao final do ciclo de fermentação, realizar o reciclo das células como parte do inóculo do próximo ciclo, reduzindo significativamente o tempo e custo de se preparar um novo inóculo (KAVANAGH & WHITTAKER, 1994). Essa estratégia, entretanto, agrava estresses impostos a leveduras, podendo causar a redução da viabilidade celular (BASSO et al., 2008). Por conta disso, é de interesse industrial que as linhagens de *S. cerevisiae* utilizadas nesse processo tenham uma tolerância alta ao etanol. Até o final da década de 1980, linhagens modificadas em laboratório eram procuradas na tentativa de se aumentar a eficiência do processo fermentativo (RODRIGUES & ANDRIETTA, 1995), porém estes organismos não sobreviviam ao processo completo. Linhagens de *S. cerevisiae* selvagens, que habitam naturalmente a cana-de-açúcar, demonstram maior resistência ao processo fermentativo, e mesmo após a reutilização dos inóculos em mais de um ciclo fermentativo conseguem se manter viáveis (BASSO et al., 1993). Dentre as várias linhagens de leveduras selvagens utilizadas atualmente pela indústria de produção de etanol brasileira, a linhagem Pedra 2 (PE-2) se destaca por diversos fatores, entre eles a alta tolerância ao etanol e alguns metabólicos tóxicos, ao estresse oxidativo e mesmo à variações de temperatura (PEREIRA et al., 2010). A resposta generalizada das leveduras ao estresse é bastante complexa, e envolve aspectos como recepção e transdução de sinais, controle transcricional e pós-transcricional, o acúmulo de proteínas protetoras e um aumento na atividade de funções de reparo (MAGER & FERREIRA., 1993), porém comparar a atividade gênica da resposta específica ao estresse de etanol da linhagem PE-2 com outras linhagens de *S. cerevisiae* pode gerar novos conhecimentos sobre os mecanismos de resistência ao etanol adquiridos por PE-2.

Em um ambiente com uma alta concentração de etanol, podem ser observados alguns efeitos comuns de estresse sobre *Saccharomyces cerevisiae*. O etanol inibe a levedura de

forma que em baixas concentrações ocorre uma diminuição da taxa de divisão e crescimento celular. Em altas concentrações, ocorre um aumento generalizado da taxa de morte celular (BIRCH & WALKER, 2000). Um ambiente com etanol faz também com que a fluidez da membrana celular das leveduras aumente, e que a integridade da membrana diminua. Como resposta ao estresse gerado pelo etanol, proteínas de choque térmico (HSP) são produzidas, diminuindo a taxa de acúmulo de RNA e proteínas, aumentando a taxa de mutações no DNA, desnaturando proteínas e enzimas e reduzindo a atividade delas (HU et al, 2007). Levantamentos e revisões realizados no passado apontam as principais respostas conhecidas ao estresse de etanol por *S. cerevisiae*, e quais genes apresentam uma resposta de regulação nesse caso específico (MA & LIU, 2010; STANLEY et al., 2009). Um dos primeiros estudos de expressão gênica durante estresse de etanol conduzido em *S. cerevisiae* mutantes (OGAWA et al., 2000) revelou uma grande regulação positiva dos genes CTT1 (codifica a catalase citosólica T), GPD1 (codifica a glicerol-3-fosfato dehidrogenase), SPI1 (codifica uma proteína da parede celular), HSP12 (codifica uma HSP associada a membrana) e HOR7 (codifica uma proteína de membrana do tipo I). Estudos realizados com o método de “*single gene knockout*” (SKGO) demonstraram que diferentes genes estão envolvidos na resposta de leveduras mutantes a diferentes concentrações de etanol (8% e 11% v/v) (KUBOTA et al, 2004). Outro estudo observou que apenas 7% dos genes regulados positivamente em condições de estresse de etanol eram os mesmos para 3 horas e 1 hora de estresse (CHANDLER et al., 2004), logo o tempo que as leveduras permanecem no meio estressante também é um fator a ser levado em conta. Uma revisão publicada em 2009 por STANLEY et al., mostrou que o resultado de quatro estudos similares (YOSHIKAWA et al., 2009; FUJITA et al., 2006; KUBOTA et al., 2004; VAN VOORST et al., 2006), porém distintos, observando mutantes sensíveis ao etanol por SKGO mostrou que apenas 18% dos genes deletados resultavam em leveduras sensíveis em pelo menos dois dos estudos, isto é, 82% dos genes deletados resultaram em sensibilidade apenas em seu estudo de origem, sendo que apenas os genes VPS36 e SMI1 tiveram resultados positivos nos quatro estudos. Outro resultado comum entre os quatro estudos foi a grande significância da síntese de aminoácidos e funções do vacúolo e mitocôndrias na tolerância ao etanol, o que foi evidenciado por uma análise dos genes em categorias GO. A impossibilidade de uma comparação apropriada dos quatro estudos corrobora com a ideia de que é necessário ter cuidado na hora de se realizar interpretações generalizadas baseadas em apenas um estudo, já que uma diferença na linhagem de levedura escolhida, ou no tempo de estresse aplicado, geram resultados com bastante disparidade (STANLEY et al., 2009).

2. Objetivos

Esse trabalho tem como objetivo principal realizar uma análise transcriptômica de leveduras da espécie *Sacchararomyces cerevisiae* (Linhagem Pedra-2) utilizando ferramentas de bioinformática para, a partir da contagem de mRNA, observar a expressão gênica diferencial de grupos submetidos a tratamentos com diferentes níveis de estresse de etanol (0% vs. 6% v/v de etanol). Além disso, os resultados serão comparados com outros estudos anteriores na área, tendo ainda como objetivo diferenciar a resposta ao estresse de etanol de Pedra-2 de outras linhagens de *S. cerevisiae*.

3. Materiais e Métodos

3.1 Cultivo e obtenção do transcriptoma de *S. cerevisiae* PE-2

Primeiramente se conseguiu uma linhagem diploide de *S. cerevisiae* PE-2, obtida do estoque do professor Luiz Carlos Bassi (ESALQ). Dessa linhagem diploide se obtiveram os progenitores haploides H3 e H4. A partir do cruzamento de H3 e H4 se formou uma linhagem diploide, cultivada em condições de crescimento normal e com 6% v/v de etanol. O material foi crescido *overnight* e inoculado (densidade ótica de cerca de 0,1) em 3 réplicas de dois meios de cultura diferentes, um sem álcool e outro 6% v/v de etanol. O material foi recolhido na fase exponencial de crescimento (OD de 0,6-0,9), peletado e congelado em nitrogênio líquido. Se extraiu o RNA total utilizando o kit RNAeasy (QUIAGEN), seguido da extração do mRNA. Foram preparadas bibliotecas de RNA-seq com o kit TruSeq Stranded mRNA Sample Prep e sequenciadas na plataforma Illumina MiSeq (2 x 75 pb).

3.2 Filtragem e mapeamento de sequencias contra modelos gênicos

Finalizado o sequenciamentos das bibliotecas por RNA-Seq, as respectivas sequências foram obtidas via internet da nuvem BaseSpace (Illumina) em arquivos diferentes para cada amostra, todos no formato “.fastq”. Para cada leitura de sequenciamento de cada amostra, foi realizado um processo de filtragem de qualidade, em que bases com um *Phred score* com probabilidade de erro maior do que 0,05 foram cortadas das sequências. Utilizando o algoritmo de mapeamento “*RNA-Seq Analysis*”, disponível na plataforma CLC *Genomic Workbench* (CLC Bio, QUIAGEN), as leituras de sequenciamento de cada amostra foram

alinhas contra um modelo gênico de *Saccharomyces cerevisiae* linhagem PE-2. Este modelo foi gerado a partir da obtenção do genoma do progenitor haploide H3, sendo que a anotação dos genes foi feita parcialmente por resequenciamento contra o genoma de *S. cerevisiae* S288c, já integralmente sequenciado, e manualmente para certos genes. Esse algoritmo realizou também uma contagem do número de vezes que cada gene se encontra representado na amostra. Os seguintes parâmetros foram utilizados no algoritmo “RNA-Seq Analysis” para estes alinhamentos:

Mismatch cost: 2

Insertion cost: 3

Deletion cost: 3

Length fraction: 0,8

Similarity fraction: 0,8

Strand specific: Both

Maximum number of hits for a read: 10

3.3 Estudo de Expressão Gênica Diferencial

Para se realizar o estudo de expressão gênica diferencial, os dados obtidos a partir da quantificação dos transcritos nas amostras foram utilizados junto do conjunto de testes estatísticos DESeq2 (LOVE et al, 2014) no software R, que gerou uma tabela comparativa entre as contagens de transcritos das duas condições de crescimento da levedura (meio com 0% e 6% v/v de etanol). A tabela obtida apresentou cálculos estatísticos de *p-value*, *fold change* e *false discovery rate* (FDR), entre outros. O critério utilizado para filtrar dados potencialmente inconsistentes dessa tabela foi estabelecido como um FDR < 0,001. Além disso foram filtrados da lista genes que não contivessem 10 contagens de *reads* em pelo menos uma das seis amostras analisadas. Sendo este estudo em que o tratamento com 0% v/v de etanol é considerado o controle, então os genes cuja expressão se encontra aumentada na condição de tratamento de etanol 6% v/v são referidos como positivamente expressos (*upregulated*). Por outro lado, genes cuja expressão se encontra diminuída na condição de estresse de etanol a 6% v/v são referidos como negativamente expressos (*downregulated*).

3.4 Análise GO

O nome dos genes que integram a lista gerada pelo DESeq2, após filtragem, foram separados em duas listas baseadas em seu *fold change*, sendo uma lista composta pelos genes regulados positivamente (214 genes) e outra pelos genes regulados negativamente (355 genes). Ambas as listas de genes foram utilizadas na ferramenta Gene Ontology Term Finder, disponível no website da SDG (Saccharomyces Genome Database, em www.yeastgenome.org) (CHERRY et al, 2012), e geraram novas listas contendo a classificação dos genes em função molecular, processo metabólico e componente celular onde se encontram. As classificações e subclassificações dadas aos genes não são mutuamente exclusivas, portanto um gene ter mais de uma classificação.

4. Resultados

4.1 Estudo de expressão gênica diferencial

O processo de mapeamento encontrou sequências de mRNAs relativas a 5722 genes diferentes no total das seis amostras. Apesar de bastante robustas, as anotações do modelo gênico de *S. cerevisiae* PE-2 utilizado no mapeamento não estavam completas, portanto deve-se levar em consideração que alguns genes podem ter sido ignorados durante o processo. Após a obtenção da tabela de resultados estatísticos obtida pelo DESeq2, e filtragem dos resultados com FDR menor que 0,001, foram mantidos em uma nova tabela os resultados de 642 genes, sendo que 355 dos genes foram encontrados como regulados negativamente e 214 regulados positivamente. A maior parte dos genes que apresentaram expressão diferencial demonstraram um *fold change* bastante próximo de 1,0 (para os positivamente expressos) e -1,0 (para os negativamente expressos), tendo poucos valores discrepantes no geral (Figura 1). Uma análise PCA (Principal Component Analysis), gerada com os dados de cada amostra, demonstrou correlação nas amostras do grupo controle e do tratamento com etanol entre elas mesmas (Figura 3). Dos genes regulados negativamente, IMD2, YLL053C, ZRT1, BNA2 e IRC18 tiveram o *fold change* mais negativo, indicando que foram os cinco genes com a expressão mais reduzida no tratamento de 6% v/v de etanol. Os cinco genes regulados mais positivamente foram FMP43, YPR036W-A, UBX6, TDH1 e TAT1 (Tabela 1).

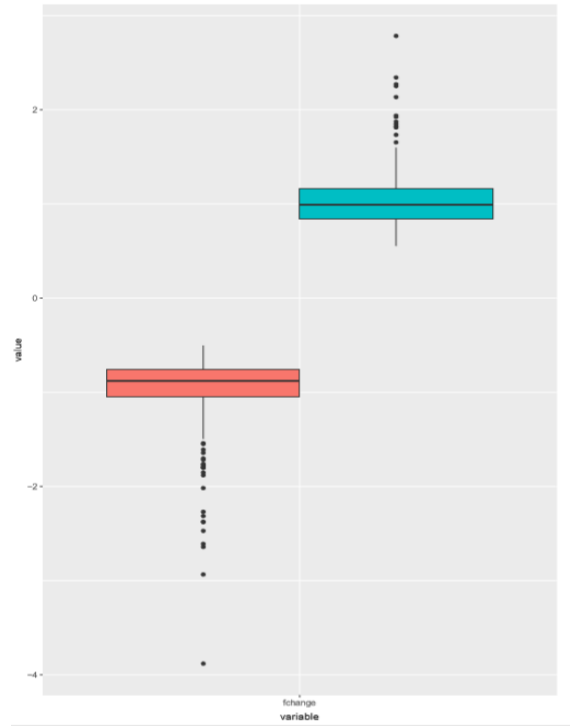


Figura 1 – Box plot formado pelos dados obtidos com o DESeq2. O eixo y representa o *fold change* (\log_2) das amostras, sendo a cor vermelha representante dos genes regulados negativamente, e azul os genes regulados positivamente.

Gene	Fold Change	Descrição
IMD2	-11,45	IMP Desidrogenase - Biossíntese de GTP
YLL053C	-4,73	Aquaporina expressa apenas em células em crescimento
ZRT1	-4,25	Transportador de Zinco pela membrana plasmática
BNA2	-4,01	Triptofano 2,3-desidrogenase putativa- Biossíntese de NAD
IRC18	-3,96	Proteína envolvida na formação da parede celular externa
FMP43	7,13	Subunidade de carreador de piruvato mitocondrial
YPR036W-A	6,36	Proteína hipotética
UBX6	5,11	Proteína reguladora de ubiquitina X
TDH1	4,89	GAPDH - Envolvida na glicólise e gliconeogênese
TAT1	4,82	Transportador de aminoácidos

Tabela 1 - Resultados estatísticos obtida a partir do DESeq2, contendo os 5 genes com as maiores diferenças na taxa de expressão (*Fold Change*) regulados negativamente e positivamente, respectivamente.

Com as anotações designadas aos genes, foi possível, a partir da tabela filtrada de 642 genes, separa-los em algumas categorias específicas, dependendo da rota metabólica que integrem, ou componente que façam parte. Diferente da análise GO, essa separação foi feita manualmente, selecionando os grupos que tiveram diferença de expressão mais óbvia, entrando nas categorias de biossíntese de cisteína (15 genes), biossíntese de lisina (2 genes), glicolização de proteínas (4 genes), biossíntese de NAD (9 genes), biossíntese de purinas (12 genes), biogênese de ribossomos (44 genes), relacionados a ribossomos (132 genes), tradução de proteínas (15 genes) e relacionados a ribossomos mitocondriais (61 genes) (Figura 2).

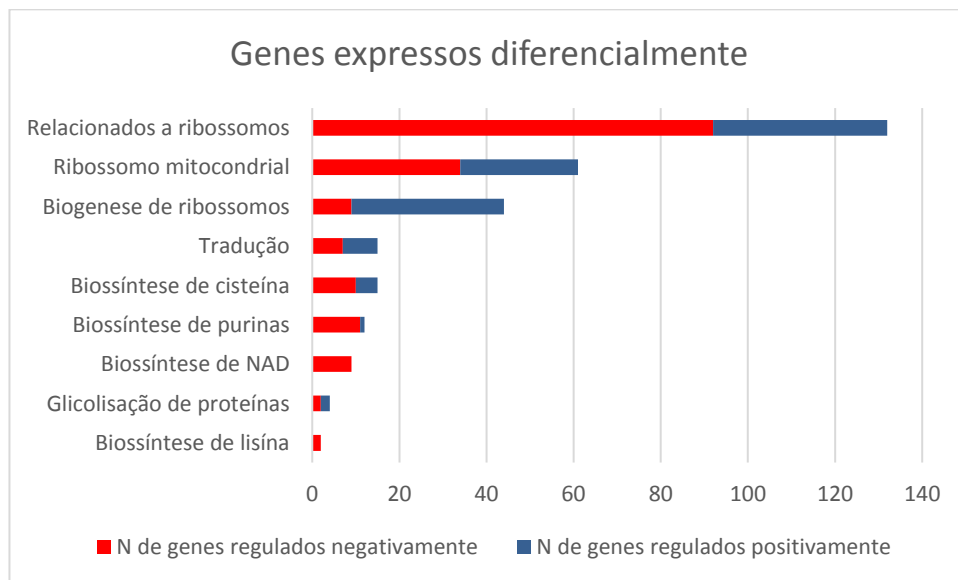


Figura 2 – Número de genes expressos diferencialmente de forma positiva e negativa no tratamento de 6% v/v de etanol, relativo ao total de cada categoria de genes.

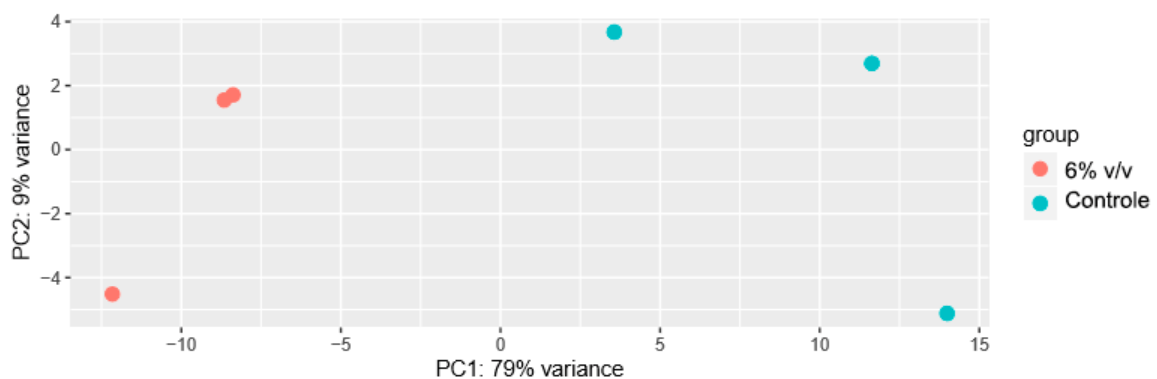


Figura 3 – Análise de componente principal (PCA). O componente principal utilizado foi a contagem de mRNAs em cada gene (após normalização).

4.2 Análise GO

Foram utilizadas separadamente as listas dos genes regulados negativamente e positivamente, e obtidas novas listas para cada categoria GO. Na classificação de componente celular, os genes regulados negativamente foram divididos em 24 subgrupos, com 93,2% desses genes fazendo parte de uma célula (GO:0044464) (classificação mais abrangente) e 20,3% fazendo parte de ribossomos (GO:0005840). Ainda em componente celular, os genes positivamente regulados foram divididos em 11 subgrupos, não apresentando genes relacionados a ribossomos, porém com 9% relacionados a pré-ribossomos (GO:0030684). Outra subdivisão presente apenas nos genes *upregulated* foi a de genes relacionados ao nucléolo (GO:0005730), representando 13,7% da amostra de 214 genes (Figura 4).

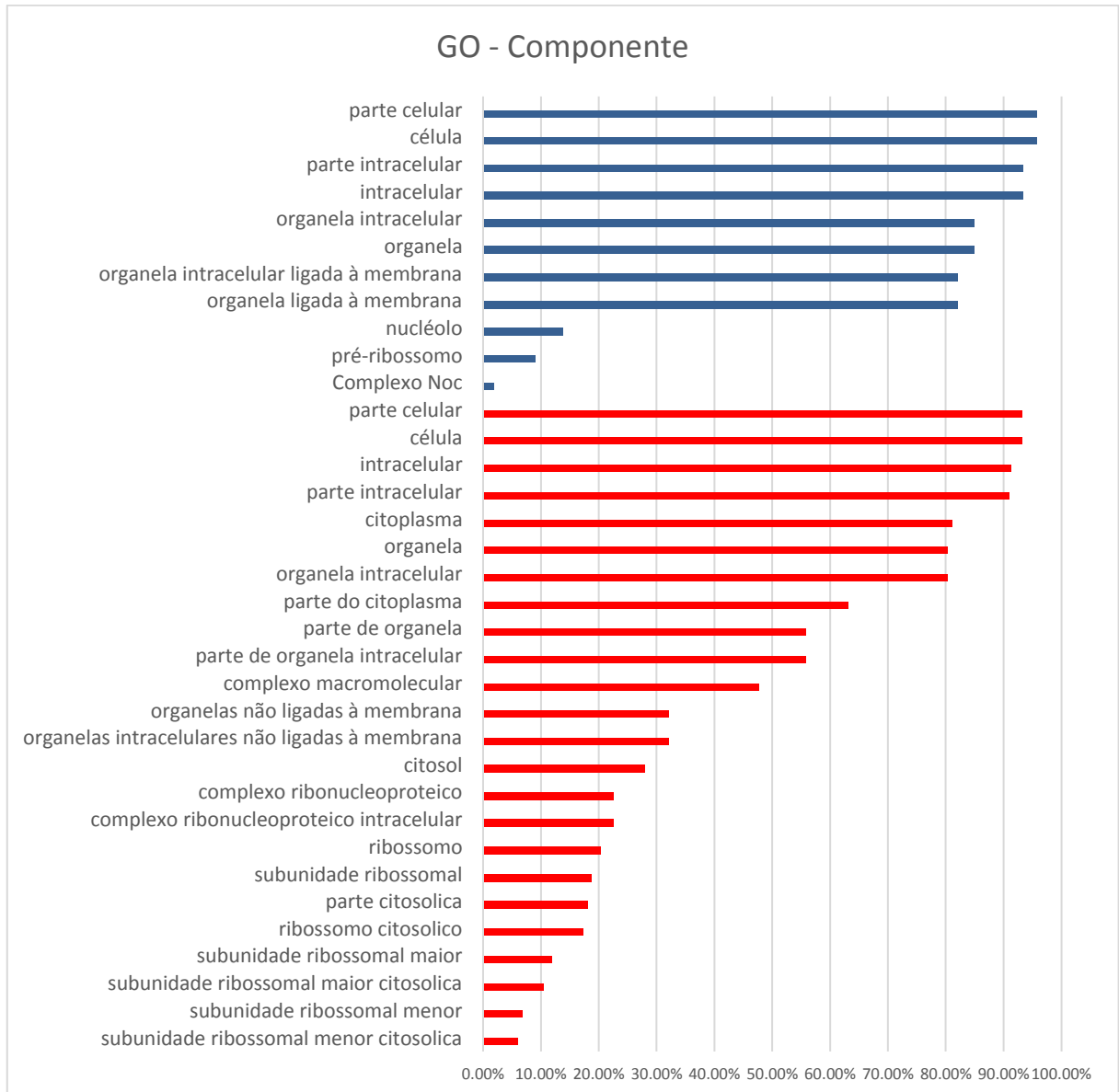


Figura 4 – Classificação dos genes baseada em componentes após análise GO. As barras azuis representam genes regulados positivamente, e as barras vermelhas genes regulados negativamente. A extensão das barras representa a porcentagem de genes da amostra total que receberam cada classificação.

Na classificação GO de função molecular os genes negativamente regulados foram classificados em “constituente estrutural de ribossomos” (GO:0003735) (18,6%) e “atividade molecular estrutural” (GO:0005198) (21,1%), enquanto os genes positivamente regulados entraram apenas na classificação de “ligação ao mRNA” (GO:0003729) (9,0%). Em ambos os casos a maior parte dos genes não foi classificado como possuindo alguma função molecular (Figura 5).

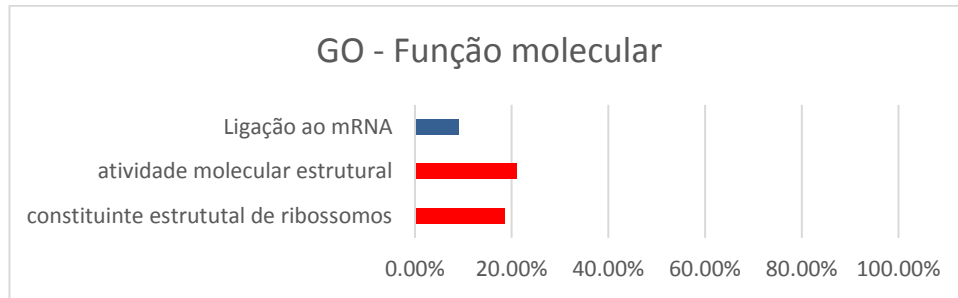


Figura 5 – Classificação dos genes baseada em função molecular. A cor azul representa genes positivamente expressos e vermelho genes negativamente expressos. A extensão das barras representa a porcentagem de genes da amostra total que receberam cada classificação.

A classificação em processo biológico gerou resultados apenas para os genes expressos negativamente, classificando-os em 14 processos biológicos (Figura 6).

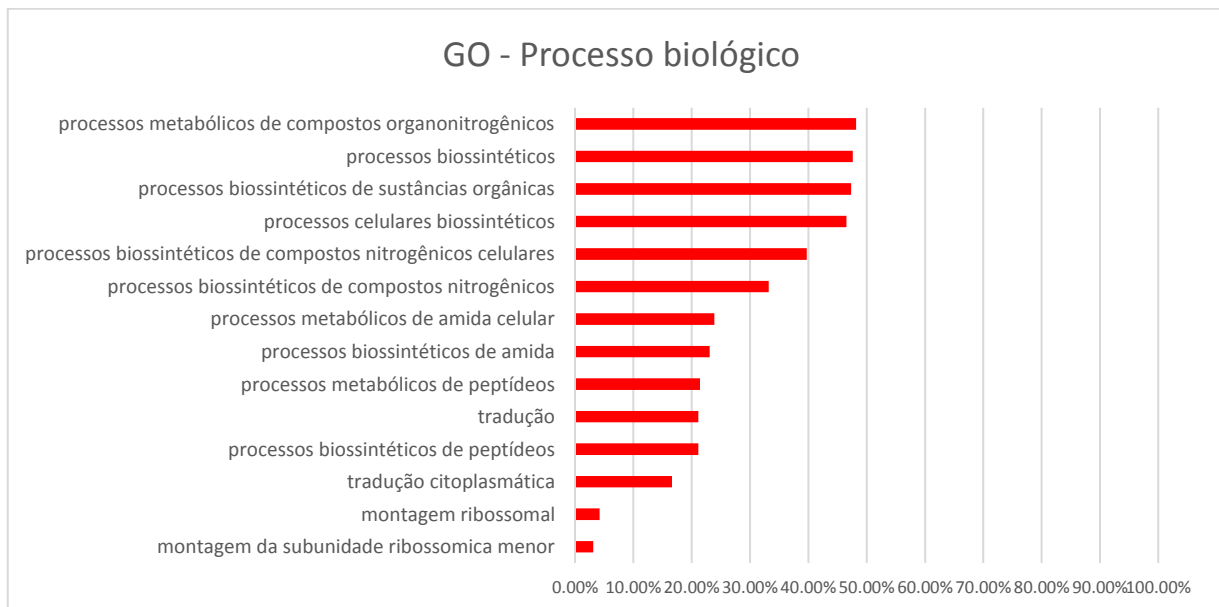


Figura 6 – Classificação dos genes baseada em processo biológico. Apenas genes negativamente expressos receberam essa classificação.

5. Discussão

Em um contexto geral, um pouco mais que a metade dos genes encontrados com taxas de regulação alteradas eram regulados negativamente (55,3%). Um maior número de genes regulados negativamente já era esperado, já que quando leveduras se encontram em condições de estresse, é comum ocorrer uma repressão global da quantidade de ribossomos na célula, e

consequentemente da síntese de proteínas. Essa redução tende a garantir um menor uso de energia que seria usada na criação de novas proteínas e no sistema de controle de qualidade de proteínas (HOLCIK & SONENBERG, 2005). A resposta ao estresse de etanol por *Saccharomyces cerevisiae* PE-2 não foge ao esperado nesse aspecto, tendo em vista também que durante este experimento, as leveduras tratadas com 6% v/v de etanol levaram um tempo maior para atingir a fase exponencial de seu crescimento *overnight* do que as leveduras do tratamento controle. De acordo com a análise GO de função molecular, aproximadamente 20% dos genes regulados negativamente tem produtos que são constituintes estruturais dos ribossomos, porém na análise manual da categorização dos genes (Figura 2), foram encontrados 92 genes regulados negativamente relacionados a ribossomos no geral, isto é, 26% desses genes. Na análise GO de processos metabólicos (Figura 6), apenas genes regulados negativamente foram categorizados, o que pode indicar uma falta de consistência na regulação positiva, e se tratar de genes regulados positivamente ao acaso, porém a análise PCA (Figura 3) não corrobora com esta ideia. Genes relacionados particularmente à biogênese de ribossomos demonstraram, no geral, aumento em expressão, representando 16% dos genes regulados positivamente. Sfp1p é um fator de transcrição relacionado à biogênese de ribossomos, e como descrito por FINGERMAN et al. (2003), o gene SFP1 tende a ser positivamente expresso em *S. cerevisiae* durante o estresse. Esse não foi o caso de PE-2, que manteve a expressão de SFP1 praticamente inalterada, porém, realizando uma análise de fatores de transcrição pela base de dados *YEASTRACT* (TEIXEIRA et al., 2018), se observou que 74% dos genes regulados positivamente eram regulados pelo fator de transcrição (TF) Sfp1. Em contrapartida, 79,7% dos genes regulados negativamente também eram regulados por Sfp1, o que gera incerteza quanto à natureza desse TF na condição de estresse por etanol.

Vários mecanismos e genes relacionados à tolerância ao estresse alcoólico por *Saccharomyces cerevisiae* já foram descritos na literatura. Muitos desses genes positivamente regulados durante o estresse estão envolvidos na composição da membrana plasmática e parede celular, porém a maior parte dos genes relacionados a membranas e paredes são regulados negativamente durante o estresse de etanol, sugerindo que possa haver um mecanismo de remodelagem da parede como resposta ao etanol (KUBOTA et al, 2004; FUJITA et al, 2006; VAN VOORST et al, 2006). Comparando os genes ativados no tratamento de PE-2 com 6% v/v de etanol, pôde ser observada a regulação positiva de alguns genes já descritos na literatura. O gene TSC13, relacionado à alongação de correntes de ácidos-graxos saturados, demonstrou ter sido regulado positivamente (*fold change*: 2,94).

HAC1, envolvido na regulação de lipídios e na biogênese da membrana, MNN10, componente do complexo maniltransferase presente na membrana plasmática, e KRE6, codificador da β -glucano sintase que atua na construção do β -glucano da parede celular, apresentaram *fold change* de 2,59, 2,58 e 2,51 respectivamente. O gene ERG6, envolvido na biossíntese de ergosterol, quando deletado em mutantes, gera indivíduos intolerantes ao etanol (KUBOTA et al, 2004; FUJITA et al, 2006; VAN VOORST et al, 2006; YOSHIKAWA et al, 2009), porém esse gene mostrou um *fold change* de -2,3, ou seja, estava regulado negativamente, assim como o esperado de todos os genes relacionados à biossíntese de ergosterol (ALEXANDER et al, 2001; CHANDLER et al, 2004), o que demonstra que nem sempre genes com produtos essenciais para a tolerância ao etanol são regulados positivamente.

Uma outra resposta comumente encontrada em *S. cerevisiae* ao estresse alcoólico é a regulação positiva de genes que codificam proteínas de choque térmico (HSP). Nesse caso não ocorreu regulação positiva relevante de nenhum gene do grupo HSP, como seria o esperado, porém o gene HDA1, que induz a produção de Hsp90, uma proteína de choque térmico responsável pela resistência a certas drogas em fungos (ROBBINS et al, 2012), teve um *fold change* de 3,39. O papel das HSPs geralmente é o mesmo de chaperonas, garantindo que proteínas mantenham uma conformação funcional, portanto genes que codifiquem chaperonas no geral podem ter o mesmo efeito durante o estresse de etanol. O etanol possui a capacidade de desnaturar proteínas, criando um acúmulo delas dentro da células, o que faz com que seja necessário que leveduras tolerantes ao etanol possuam um bom sistema de reparo de proteínas (MA & LIU, 2010). PE-2 demonstrou um aumento na expressão de JJJ1 (*fold change*: 2,26), uma co-chaperona relacionada com a biogênese de ribossomos. O acúmulo de trealose parece ser bastante comum em leveduras em estresse de etanol, e células que não conseguem acumulá-la tendem a ter seu crescimento retardado (OGAWA et al, 2000; KAINO & TAKAGI, 2008), porém PE-2, mesmo em condições de estresse de etanol, não demonstrou nenhum aumento significativo na regulação dos genes de produção de trealose, mas também não apresentou expressão aumentada dos genes degradadores de trealose, que também, em outros experimentos, tenderam a ter um aumento de expressão (CHANDLER et al, 2004).

De acordo com estudos realizados por ALEXANDRE et al. (2001) e CHANDLER et al. (2004), a maior parte dos genes de *S. cerevisiae* relacionados com a regeneração de

NADH/NADPH são regulados positivamente em situações de estresse alcoólico, mas não foi isso que se observou em PE-2. Com exceção de TDH1, relacionado com a glicólise, todos os outros genes envolvidos na regeneração NADH/NADPH não demonstraram aumento de expressão, e alguns genes envolvidos na biossíntese de NAD, como os genes da família BNA (biossíntese de NAD a partir do triptofano), demonstraram uma diminuição em sua taxa de expressão. Foi encontrado por MA & LIU (2010) que o controle (selvagem) demonstrava uma queda na expressão de genes relacionados à regeneração de NADH/NADPH após 6 horas de experimento, porém a linhagem utilizada não conseguiu estabelecer uma cultura, portanto podendo significar que uma diminuição da expressão por conta de tempo em estresse não foi o que ocorreu com PE-2. Ainda de acordo com MA & LIU (2010), a observação que fazem após a análise dos dados de diversos artigos envolvendo a transcriptômica da tolerância ao etanol em *S. cerevisiae*, é de que a resposta ao estresse observada nem sempre demonstra características de tolerância, e que apenas características funcionais e a própria demonstração da tolerância podem ser levadas em conta para se prever a performance de uma linhagem à fermentação alcoólica.

TEIXEIRA et al. (2006) e MA & LIU (2010) demonstraram que grande parte dos genes regulados positivamente durante o estresse de etanol eram regulados por principalmente 4 fatores de transcrição: Msn2p, Msn4p, Yap1p e Hsf1p. Esses TFs foram analisados utilizando a ferramenta YEASTRACT junto das listas de genes positivamente e negativamente regulados. Msn2p e Msn4p, indutores de genes de estresse ambiental por elementos de resposta ao estresse (STRE), não demonstraram resultados confiáveis (p -value baixo), e genes *downregulated* e *upregulated* tiveram uma mesma porcentagem de genes regulados por tais TFs. Yap1p, um TF requerido na tolerância ao estresse oxidativo, também não demonstrou dados conclusivos. Hsf1p foi o único fator de transcrição sugerido anteriormente que apresentou valores bem definidos entre os genes regulados positiva e negativamente. Aproximadamente 43% dos genes *downregulated* são regulados por Hsf1p, com um p -value menor que $1,00E-15$, enquanto os genes *upregulated* demonstraram 29,2% dos genes regulados por Hsf1p, porém com um p -value de 0,03, ou seja, pouco confiável. Esse resultado vai contra o esperado pela literatura anterior, já que Hsf1p se trata de um TF envolvido na transcrição de genes relacionados ao dobramento de proteínas, detoxificação, geração de energia e organização da parede celular (HAHN et al., 2004; EASTMOND & NELSON, 2006).

6. Conclusão

A resposta ao estresse de etanol foi observada em *Saccharomyces cerevisiae* PE-2 quando comparando um grupo tratado com 6% v/v de etanol a um grupo controle. Essa resposta ficou evidenciada pela regulação negativa encontrada em diversos genes, principalmente dos genes relacionados a ribossomos e a produção e reparo de proteínas, o que resultou em um tempo de crescimento mais lento das leveduras no meio com etanol até a fase exponencial. Apesar de *S. cerevisiae* demonstrar tolerância ao etanol, o mecanismo por qual essa tolerância é expressa não pode ser observado com clareza, e inclusive muitos dos dados obtidos foram contra trabalhos anteriores com outras linhagens. Esses resultados podem indicar que os mecanismos de tolerância ao etanol de PE-2 não são os mesmos dos convencionais, porém podem também ter sido fruto de um ainda baixo conhecimento da resposta específica ao estresse de etanol por leveduras, aliado a uma falta de padrão dos trabalhos realizados nessa área.

7. Referências Bibliográficas

BROWN, T.A., **Genomes**. 2002. 2. ed. Oxford: Wiley-Liss.

WANG, Z.; GERSTEIN, M.; SNYDER, M. **RNA-Seq: a revolutionary tool for transcriptomics**. 2009. *Nature Reviews Genetics*. 10(1): 57 – 63.

OKONIEWSKI, M.J.; MILLER, C.J. **Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations**. 2006. *BMC Bioinformatics*. 7: 276.

ROYCE, T.E.; ROZOWSKY, J.S.; GERSTEIN, M.B. **Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification**. 2007. *Nucleic Acids Research*. 35(15): e99.

MARIONI, J.; MASON, C.; MANE, S.; STEPHENS, M.; GILAD, Y. **RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays**. 2008. *Genome Research*. 18(9): 1509 – 17.

SONESON, C.; DELORENZI, M. **A comparison of methods for differential expression analysis of RNA-Seq data**. 2013. *BMC Bioinformatics*. 14: 91.

OSHLACK, A.; ROBINSON, M.D.; YOUNG, M.D. **From RNA-Seq reads to differential expression results**. 2010. *Genome Biology*. 11: 220.

BULLARD, J.H.; PURDOM, E.; HANSEN, K.D.; DUDOIT, S. **Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments**. 2010. *BMC Bioinformatics*. 11: 94.

LOVE, M.I.; HUBER, W.; ANDERS, S. **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. 2014. *Genome Biology*. 15: 550.

THE GENE ONTOLOGY CONSORTIUM. **Gene Ontology: tool for the unification of biology**. 2000. *Nature Genetics*. 25 (1): 25 – 9.

KAVANAGH, K.; WHITTAKER, P. **Application of the Melle-Boinot process to the fermentation of xylose by *Pachysolen tannophilus***. 1994. *Applied Microbiology and Biotechnology*. 42(1): 28 – 31.

BASSO, L.C.; DE AMORIM, H.V.; DE OLIVEIRA, A.J.; LOPES, M.L. **Yeast selection for fuel ethanol production in Brazil**. 2008. *FEMS Yeast Research*. 8(7): 1155 – 63.

RODRIGUES, M.I.; ANDRIETTA, M.G.S. **Controle da fermentação alcoólica através de testes microbiológicos e bioquímicos**. 1995. Apostila. p. 54.

BASSO, L.C.; OLIVEIRA, A.J.; ORELLI, A.A.; CAMPOS, C.R.; GALLO, C.R.; AMORIM, H.V.; **Dominância das leveduras contaminantes sobre as linhagens industriais avaliada pela técnica de cariotipagem.** 1993. Anais Congresso Nacional da STAB, Piracicaba, v.5, n.1, p.5.

PEREIRA, F.B.; GUIMARÃES, P.M.R.; TEIXEIRA, J.A.; DOMINGUES, L. **Selection of *Saccharomyces cerevisiae* strains for efficient very high gravity bio-ethanol fermentation processes.** 2010. Biotechnology Letters. 32(11): 1655 – 61.

MAGER, W.H.; FERREIRA, P.M. **Stress Response of Yeast.** 1993. Biochemical Journal. 290(1): 1 – 13.

BIRCH, M.R.; WALKER, M.G. **Influence of magnesium ions on heat shock and ethanol stress responses of *Saccharomyces cerevisiae*.** 2000. Enzyme and Microbial Technology. 26 (9 - 10): 678 – 687.

HU, X.H.; WANG, M.H.; TAN, T.; LI, J.R.; YANG, H.; LEACH, L.; ZHANG, R.M.; LUO, Z.W. **Genetic Dissection of Ethanol Tolerance in the Budding Yeast *Saccharomyces cerevisiae*.** 2007. Genetics. 175(3): 1479 – 1487.

MA, M.; LIU, Z.L. **Mechanisms of ethanol tolerance in *Saccharomyces cerevisiae*.** 2010 Applied Microbiology and Biotechnology. 87(3): 829 – 45.

STANLEY, D.; BANDARA, A.; FRASER, S.; CHAMBERS, P.J.; STANLEY, G.A. **The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*.** 2009. Journal of Applied Microbiology. 109(1): 13 – 24.

OGAWA, Y.; NITTA, A.; UCHIYAMA, H.; IMAMURA, T.; SHIMOI, H.; ITO, K. **Tolerance mechanism of the ethanol-tolerant mutant of sake yeast.** 2000. Journal of Bioscience and Bioengineering. 90(3): 313 – 320.

KUBOTA, S.; TAKEO, I.; KUME, K.; KANAI, M.; SHITAMUKAI, A.; MIZUNUMA, M.; MIYAKAWA, T.; SHIMOI, H.; IEFUJI, H.; HIRATA, D. **Effect of Ethanol on Cell Growth of Budding Yeast: Genes That Are Important for Cell Growth in the Presence of Ethanol.** 2004. Biosciences, Biotechnology, and Biochemistry. 68(4): 968 – 972.

CHANDLER, M.; STANLEY, G.A.; ROGERS, P.; CHAMBERS, P. **A genomic approach to defining the ethanol stress response in the yeast *Saccharomyces cerevisiae*.** 2004. Annals of Microbiology. 54(4): 427 – 454.

YOSHIKAWA K.; TANAKA, T.; FURUSAWA, C.; NAGAHISA, K.; HIRASAWA, T.; SHIMIZU, H. **Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in *Saccharomyces cerevisiae*.** 2009. FEMS Yeast Research. 9(1): 32 – 44.

FUJITA K.; MATSUYAMA, A.; YOSHINORI, K.; IWAHASHI, H. **The genome-wide screening of yeast deletion mutants to identify the genes required for tolerance to ethanol and other alcohols.** 2006. FEMS Yeast Research. 6(5): 744 – 50.

VAN VOORST, F.; HOUGHTON-LARSEN, J.; JONSON, L.; KIELLAND-BRANDT, M.C.; BRANDT, A. **Genome-wide identification of genes required for growth of *Saccharomyces cerevisiae* under ethanol stress.** 2006. *Yeast*. 23(5): 351 – 9.

CHERRY, J.M.; HONG, E.L.; AMUNDSEN, C.; BALAKRISHNAN, R.; BINKLEY, G.; CHAN, E.T.; CHRISTIE, K.R.; COSTANZO, M.C.; DWIGHT, S.S.; ENGEL, S.R.; FISK, D.G.; HIRSCHMAN, J.E.; HITZ, B.C.; KARRA, K.; KRIEGER, C.J.; MIYASATO, S.R.; NASH, R.S.; PARK, J.; SKRZYPEK, M.S.; SIMISON, M.; WENG, S.; WONG, E.D. **Saccharomyces Genome Database: the genomics resource of budding yeast.** 2012. *Nucleic Acids Research*. 40: D700-5.

HOLCIK, M.; SONENBERG, N. **Translational control in stress and apoptosis.** 2005. *Nature Reviews Molecular Cell Biology*. 6: 318-327.

FINGERMAN, I.; NAGARAJ, V.; NORRIS, D.; VERSHON, A.K. **Sfp1 plays a key role in yeast ribosome biogenesis.** 2003. *Eukaryot Cell*. 2: 1061 – 1068.

TEIXEIRA, M.C.; MONTEIRO, P.T.; PALMA, M.; COSTA, C.; GODINHO, C.P.; PAIS, P.; CAVALHEIRO, M.; ANTUNES, M.; LEMOS, A.; PEDREIRA, T.; SÁ-CORREIA, I. **YEASTRACT, an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*.** 2018. *Nucleic Acid Research*. 46: 348-353.

ALEXANDER, H.; ANSANAY-GALEOTE, V.; DEQUIN, S.; BLONDIN, B. **Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*.** 2001. *FEBS Letters*. 498(1): 98 – 103.

ROBBINS, N.; LEACH, M.D.; COWEN, L.E. **Lysine deacetylases Hda1 and Rpd3 regulate Hsp90 function thereby governing fungal drug resistance.** 2012. *Cell Reports*. 2(4): 878 – 88.

KAINO, T.; TAKAGI, H. **Gene expression profiles and intracellular contents of stress protectants in *Saccharomyces cerevisiae* under ethanol and sorbitol stresses.** 2008. *Applied Microbiology and Biotechnology*. 79(2): 273 – 83.

ALEXANDRE, H.; ANSANAY-GALEOTE, V.; DEQUIN, S.; BLONDIN, B. **Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*.** 2001. *FEBS Letters*. 498(1): 98 – 103.

HAHN, J.S.; HU, Z.; THIELE, D.J.; IYER, V.R. **Genome-wide analysis of the biology of stress response through heat shock transcription factor.** 2004. *Molecular Cell Biology*. 24: 5249 - 5256.

EASTMOND, D.L.; NELSON, H.C. **Genome-wide analysis reveals new roles for the activation domains of the *Saccharomyces cerevisiae* heat shock transcription factor (Hsf1) during the transient heat shock response.** 2006. *Journal of Biological Chemistry*. 281: 32909 – 32921.