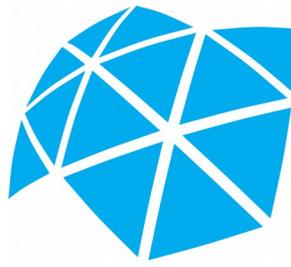


**Universidade Estadual Paulista “Júlio de Mesquita Filho”  
Instituto de Biociências de Botucatu  
Programa de Pós-Graduação em Biotecnologia**

**Identificação de assinaturas sistêmicas associadas à tolerância ao etanol em  
linhagens de *Saccharomyces cerevisiae***

**Ivan Rodrigo Wolf**

**Botucatu – SP  
2019**



**Universidade Estadual Paulista “Júlio de Mesquita Filho”  
Instituto de Biociências de Botucatu  
Programa de Pós-Graduação em Biotecnologia**

**Identificação de assinaturas sistêmicas associadas à tolerância ao etanol em  
linhagens de *Saccharomyces cerevisiae***

**Doutorando: Ivan Rodrigo Wolf  
Orientador: Dr. Guilherme Targino Valente**

Tese apresentada ao Programa de Pós-Graduação em Biotecnologia do Instituto de Biociências de Botucatu da Universidade Estadual Paulista “Júlio de Mesquita Filho”, para obtenção do título de doutor.

**Botucatu - SP  
2019**

W853i

Wolf, Ivan Rodrigo

Identificação de assinaturas sistêmicas associadas à tolerância ao etanol em linhagens de *Saccharomyces cerevisiae* / Ivan Rodrigo Wolf. -- Botucatu, 2019

118 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Botucatu  
Orientador: Guilherme Targino Valente

1. Biologia de Sistemas. 2. *Saccharomyces cerevisiae*. 3. Bioinformática. 4. Biologia celular. 5.

Tolerância ao Etanol. I. Título.

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

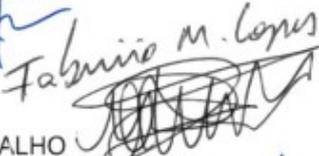
Essa ficha não pode ser modificada.

**ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE IVAN RODRIGO WOLF, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.**

Aos 01 dias do mês de julho do ano de 2019, às 14:00 horas, no(a) Sala B - Central de Aulas do IBB, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. GUILHERME TARGINO VALENTE - Orientador(a) do(a) Departamento de Engenharia de Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônômicas de Botucatu - UNESP, Prof. Dr. RAFAEL PLANA SIMÕES do(a) Departamento de Engenharia de Bioprocessos e Biotecnologia / Faculdade de Ciências Agrônômicas de Botucatu - UNESP, Prof. Dr. FABRÍCIO MARTINS LOPES do(a) Departamento Acadêmico de Computação / Universidade Tecnológica Federal do Paraná, Prof. Dr. ROBSON FRANCISCO CARVALHO do(a) Departamento de Morfologia / Instituto de Biociências de Botucatu - UNESP, Professor Associado PAULO EDUARDO MARTINS RIBOLLA do(a) Departamento de Parasitologia / Instituto de Biociências - IB - Unesp - Câmpus de Botucatu, sob a presidência do primeiro, a fim de proceder a arguição pública da TESE DE DOUTORADO de IVAN RODRIGO WOLF, intitulada **Identificação de assinaturas sistêmicas associadas à tolerância ao etanol em linhagens de *Saccharomyces cerevisiae***. Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: APROVADO \_\_\_\_\_. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

  
Prof. Dr. GUILHERME TARGINO VALENTE

Prof. Dr. RAFAEL PLANA SIMÕES 

Prof. Dr. FABRÍCIO MARTINS LOPES 

Prof. Dr. ROBSON FRANCISCO CARVALHO 

Professor Associado PAULO EDUARDO MARTINS RIBOLLA 

**“Eu sei que você acredita que entende o que acha que eu disse, mas não tenho certeza se você entende que o que você ouviu não é o que eu quis dizer.”**

- Uma das infinitas frases exibidas no Gaussian, um programa para cálculo de estrutura eletrônica.

Dedico este trabalho a meu pai, minha mãe, minha família e aos meus amigos,  
estes ajudaram a constituir os alicerces do que sou.

## Agradecimentos

Agradeço à *Deus*, pois nas horas mais escuras uma luz sempre brilhou na forma de um familiar, um amigo, o trecho de um texto ou um vídeo, um besouro de ponta cabeça na calçada... enfim, em todo e qualquer sinal de esperança. Hoje sinto sua presença naqueles que estão a minha volta.

Agradeço à *Vida* e a *Morte*, à primeira gosta de rir da nossa cara e de nos fazer tropeçar, mas sempre nos ajuda a levantar e nos empurra para seguir adiante, mais longe e mais determinados; à segunda sempre está ao nosso lado com um olhar de ternura e um sorriso reconfortante e alegre que nos lembram constantemente de darmos o nosso melhor durante essa jornada. Elas são os símbolos de início e fim que nos ensinam que passamos por diversos ciclos e que tudo tende a melhorar, terminar bem e (de preferência) em uma grande bagunça.

Agradeço aos meus pais *Ivan Wolf* e *Inês Maristela Palamar Wolf*, pela vida maravilhosa que me proporcionaram e por me ensinarem os bons princípios que me orientam fundamentalmente. Lembrar de vocês me reconforta e preenche meu coração, por isso sei que nunca vou estar sozinho. Espero conseguir passar a diante pelo menos um pouco do amor e carinho que recebi de vocês. Se os pais incomodam os filhos, só tenho a agradecer o incomodo que recebi, pois sei que são a prova do quanto vocês me amam e me querem bem. Graças a vocês não tenho medo da vida, da dor, do trabalho e do que vier pela frente. Amo vocês!

Agradeço à minha vó Jaci do Rocio Palamar, pois ao longo dos anos que se passaram depois do seu falecimento notei que o caminho que segui me levou a compreender melhor a doença que te afligi. Esse caminho também, pouco a pouco, me forneceu as ferramentas para ajudar a, quem sabe um dia, encontrar a cura que teria permitido você estar junto de nós. Se hoje estou conseguindo o título de doutor, grande influência veio dessa vontade de ter salvo você. Te amo onde quer que você esteja.

Agradeço à algumas pessoas que são como irmãos e irmãs para mim: *Jefferson Andrei Silva*, *Marina Krasniak*, *Aline Bonk*, *Camila Borão*, *Danyella Nunes*, *Joni Marcos Lopes*, *Luiz Ricardo Fantim*, *Oliver Lie Bueno*, *Everton Souza*. Por cada um de vocês tenho amor, carinho, amizade, gratidão e admiração. Mesmo depois de muitos anos terem se passado nós continuamos os mesmos. Vocês já me viram com alegria e com raiva, feliz e triste, melancólico e otimista, altruísta e egoísta, flexível e teimoso, legal e insuportável. Vocês viram o que tenho de melhor e também o que tenho de pior. Ainda assim, me recebem com um sorriso no rosto, algo para beber, conversam comigo até tarde (quando não até o raiar do sol), riem comigo e combinam o próximo encontro. Vocês sempre serão pessoas especiais para mim. Sinto cada um de vocês vivos dentro do meu coração, e assim espero viver no coração de vocês.

Agradeço aos meus primos e primas *Ederson Nogueira dos Santos*, *Tháise Nogueira dos Santos*, *Edimar Wolf* e *Edicler Wolf*. Durante muito tempo vocês foram meus irmãos e irmãs, assim não me senti um filho único. Com vocês aprendi a desenhar um pouco melhor, ler um pouco mais e a usar minha imaginação. Obrigado por todas as aventuras, acampamentos, viagens, conversas e os finais de semana brincando com Lego, jogando RPG de mesa, Xadrez, Megaman, Zelda, Mario, Mario Party, GTA, Resident Evil, Doom, Grandia, Crash Bandicoot (...) em noitadas regadas à guloseimas variadas e muitas risadas.

Agradeço ao meu orientador *Guilherme Targino Valente*, você me colocou diante de desafios que exauriram minhas forças, acabaram com minha mente, abalaram minha fé e que eu realmente achei que não fosse capaz de superar. Mas sei que você fez isso sabendo que eu estaria preparado para superar cada etapa e cada desafio. Essas superações profissionais também se devem ao fato de que você estava lá para me orientar, me ajudar, e me mostrar que com foco e competência posso resolver qualquer problema! Você me ensinou a ser um profissional melhor do que eu era. E, de quebra, se mostrou uma pessoa extremamente sábia, me deu conselhos para a vida e ainda foi humilde de me escutar quando eu tinha uma contraposição. Você é um grande professor, orientador e profissional. Obrigado!

Agradeço à *Amanda Piveta Schnepfer*, que chegou de mansinho e me cativou com seu jeito singular e um sorriso que está no seu olhar. Você me mostrou como fazer fogo com duas pedras. Por sua causa o sabor, a cor e o cheiro do vinho nunca mais serão os mesmos. Obrigado!

Agradeço à todo o pessoal do SBGL (*Systems Biology and Genomic Laboratory – FCA – UNESP de Botucatu*) e LGI (*Laboratório de Genômica Integrativa – IBB - UNESP de Botucatu*), em especial a *Lauana Fogaça de Almeida, Rodrigo de Oliveira Almeida, Luiz Henrique Cardozo, Camila Cristina de Oliveira, Lucas Lazari, Lucas Farinazzo, Guilherme Luz, Eric Kazuo Kawagoe, Diego Marques, Jordana Oliveira, Adauto Cardoso, Rafael Takahiro Nakajima, Luiz Augusto Bovolenta e Érica Ramos*. Obrigado pelas muitas horas de trabalho, pelas conversas no laboratório, pelas festas, pelas dicas, pelas idas ao bar, pela ajuda, pelos favores feitos e pelos favores cobrados. Tudo que vi, vivi e aprendi com cada um de vocês vêm a acrescentar algo positivo para melhorar a vida que tenho. Espero que possamos nos ver de novo para compartilhar experiências e histórias. Obrigado!

Agradeço aos professores *Rafael Plana Simões, César Martins, Alexandre Dal Pai, José Luiz Rybarczyk Filho* e as professoras *Rejane Maria Tommasini Grotto, Carla dos Santos Riccardi, Ivana Cesarino, Valéria Cristina Rodrigues Sarnighausen, Sarita Candida Rabelo e Lilian Cristina Pereira*. Aprendi muito com suas lições profissionais e conversas casuais. Vocês ajudaram a expandir os meus horizontes.

Agradeço aos professores da *Universidade Estadual de Londrina - UEL*. Em especial aos meus ex-orientadores *Laurival Antônio Vilas Bôas e Rogério Fernandes de Souza*, obrigado pelos desafios, por jamais duvidarem da minha competência, pelas dicas, sugestões, críticas e conselhos. Vocês são uma grande inspiração profissional.

Agradeço também aos amigos e amigas *Natália Venturelli, Marcos Edgar Herkenhoff, Raquel Bozini Gallo, Jaqueline Fernanda Dionisio, Juceli Gonzalez Gouveia, Ana Paula Scaramal Ricieto, Carla Suzuki Altrão, Carolina Galdino Gumiero Ribeiro e Rachel Colauto Milanezi Aguiar* e à todos os outros cujo os nomes não cabem aqui. Vocês são pessoas incríveis. Que nossas amizades continuem através do tempo e do espaço. Tenho certeza que não vão faltar oportunidades para um churrasco, um almoço, um café, um barzinho... Que todos vocês tenham sucesso, felicidade e alegria em suas vidas, vocês merecem!

Agradeço todos os *docentes e técnicos* do Departamento de Bioprocessos e Biotecnologia da Faculdade de Ciências Agrônômicas, UNESP- Câmpus de Botucatu, em especial à *Raquel Francisco Rubio, José Eduardo Gomes Montanha, Leonardo Nazario de Moraes, Jackson Eliezer Neves Batalha, Marcelo Felipe Matias, Martin Kássio Leme*

*da Silva, Firmo Sousa Campos.* Vocês fazem o departamento funcionar, me ajudaram, me ensinaram e me trataram como um colega de trabalho, obrigado por tudo!

Agradeço aos *graduandos* do Departamento de Bioprocessos e Biotecnologia da Faculdade de Ciências Agrônômicas, UNESP- Câmpus de Botucatu. Eu entrava na sala para ensinar, mas depois de cada curso, palestra ou aula senti que havia aprendido mais do que vocês. A lembrança dos seus rostos e das experiências que trocamos vão me manter, para sempre, um jovem aprendiz.

Agradeço à *Universidade Estadual Paulista Júlio de Mesquita Filho – UNESP - Botucatu*, pelo suporte, estrutura e bons professores que contribuíram imensamente para minha formação. E aproveito para agradecer ao *Programa de Pós-Graduação em Biotecnologia* do *Instituto de Biociências*, pela oportunidade e suporte. Aproveito também para agradecer aos *colegas da pós-graduação*, aqui seria injusto tentar citar qualquer nome; e também à secretaria da pós graduação, em especial ao *Davi Barcellos de Oliveira Müller* sempre solucionando os problemas burocráticos junto à secretaria desde o início do meu doutorado.

Agradeço à *família Sidol*, em especial à filha mais velha *Poli*. Nossas vidas agora seguem caminhos separados, mas vou levar comigo lembranças carinhosas e felizes de todos. Com vocês aprendi valiosas lições de vida. Acho que sem nossa convivência, muito provavelmente, eu não teria chegado até aqui. Obrigado pelos anos de boa companhia e amizade. Desejo à vocês tudo de bom.

Agradeço à *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)* pelo primeiro ano de auxílio e à *Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Processo número 2015/19211-7)* pelo auxílio financeiro durante esses anos de muito trabalho e aprendizado do doutorado.

## Resumo

A crescente demanda por combustíveis fósseis e as instabilidades econômicas relacionadas à sua utilização despertam o interesse no desenvolvimento de biocombustíveis tais como o bioetanol. O processo de produção mais comum do bioetanol é a tecnologia de primeira geração, no qual o organismo mais amplamente utilizado é a levedura *Saccharomyces cerevisiae*. No entanto, a alta concentração de etanol gera toxicidade para *S. cerevisiae* e constitui um dos fatores limitantes na produção deste bioetanol. Assim, a produção de linhagens mais resistentes a esse estressor constitui um ponto chave no desenvolvimento dos processos biotecnológicos relacionados ao bioetanol. Nesse contexto, análises sistêmicas são pouco empregadas para o entendimento do fenômeno de tolerância ao etanol, deixando a busca por genes candidatos algo bastante laborioso e custoso. Assim, as diferenças entre linhagens pouco tolerantes (LT) e altamente tolerantes (HT) ao etanol (EtOH) foram analisadas por meio de ferramentas de bioinformática como a biologia de sistemas e a análises de transcriptomas. Os resultados mostraram que as linhagens HT e LT utilizam inicialmente um sistema de resposta à estresse comum, mas que devido as mudanças na estrutura das redes metabólicas, regulatórias e de interação proteína-proteína de cada linhagem, mecanismos de resposta ao estresse por etanol distintos são ativados em cada grupo. As linhagens LT mantém a homeostase da célula com o ciclo celular ativo, efetuando o ajuste dos metabolismos ao custo de que elementos transponíveis ficam ativos e agentes anti oxidantes importantes deixam de ser produzidos, permitindo que ocorram danos no DNA. As linhagens HT interrompem seu ciclo celular e boa parte do metabolismo, porém, elas mantêm a produção e reparo de proteínas para estabilizar as membranas celulares e o controle de espécies reativas de oxigênio (ROS) em conjunto com a metabolização do etanol. Uma vez que o EtOH está sendo utilizado para obtenção de carbono no ciclo do ácido cítrico, as linhagens HT resistem a concentrações maiores de etanol. Assim, o conhecimento aqui gerado poderá ser aproveitado para enriquecer as discussões futuras sobre o fenômeno estudado e o desenvolvimento biotecnológico.

**Palavras-chave:** Biologia de Sistemas, *Saccharomyces cerevisiae*, Bioinformática, Biologia celular, Tolerância ao Etanol

## **Abstract**

The increasing demand and economic instabilities related to the use of fossil fuels raised interest in the development of biofuels such as bioethanol. The most common bioethanol production process is from the first generation technology, in which the most widely used organism is the yeast *Saccharomyces cerevisiae*. However, the high concentration of ethanol generates toxicity to *S. cerevisiae* and this is one of the limiting factors in the bioethanol production. Thus, the production of more resistant strains to this stressor is crucial to the development of biotechnological processes related to bioethanol. The systems analyzes are poorly explored to understand the ethanol tolerance phenomenon, leading the search for candidate genes labor intensive and expensive. Thus, the differences between low tolerant (LT) and high tolerant (HT) strains to ethanol (EtOH) were analyzed using bioinformatics tools like systems biology and transcriptome analysis. The results showed that HT and LT strains initially used a common stress response system but due to changes in the network structure of metabolic, regulatory, and protein-protein interactions of each strain, different mechanisms of stress response are activated in each group. LT strains maintain the cellular homeostasis activating cell cycle, and adjusting metabolisms, despite activation of transposable elements and disruption of important anti-oxidant agents production, allowing the occurrence of DNA damage. HT strains disrupt the cell cycle and most of metabolisms, but maintaining protein production and repair to stabilize cell membranes and control reactive oxygen species (ROS), working together with the ethanol metabolism. Moreover, HTs are using the EtOH to obtain carbon in the citric acid cycle, thus, it may be the reason of HT strains surpass high ethanol concentration. Altogether, the knowledge here generated may be used to enrich future discussions about the studied phenomenon and to provide biotechnological development.

**Keywords:** Systems Biology, *Saccharomyces cerevisiae*, Bioinformatics, Cell Biology, Ethanol Tolerance

## Lista de Ilustrações

Figura 1: Produção de bioetanol (bilhões de litros) em 2012	19
Figura 2: Classificações do bioetanol como biocombustível	20
Figura 3: Filogenia de <i>Saccharomyces cerevisiae</i>	21
Figura 4: Efeito de várias concentrações de etanol no crescimento de <i>Saccharomyces cerevisiae</i>	22
Figura 5: Nuvem de palavras mostrando termos ontológicos relacionados à tolerância ao etanol em <i>S. cerevisiae</i>	24
Figura 6: Tipos de dados e técnicas de <i>Machine learning</i>	29
Figura 7: Desenho experimental	33
Figura 8: Representação do corte feito nos dados de crescimento para ajuste ao modelo	47
Figura 9: Geração das redes de controle e tratamento	50
Figura 10: Visão geral da análise metabolômica	54
Figura 11: Análise do conjunto referente às DAMs	55
Figura 12: Comparação quantitativa e qualitativa dos genes diferencialmente expressos entre as linhagens analisadas	57
Figura 13: Diagramas comparando genes e termos GO encontrados em diferentes estudos	58
Figura 14: Via do metabolismo de arginina e prolina	65
Figura 15: Modelo proposto para o fluxo de etanol no interior da célula da levedura	66
Figura 16: Genes diferencialmente expressos na provável via de tamponamento de etanol	67
Figura 17: Ensaio de viabilidade celular e contagem de células que acumulam ROS	68
Figura 18: Distribuição dos degrees para a rede controle e rede tratamento	72
Figura 19: Análise da função DDC normalizada para redes de controle e tratamento	73
Figura 20: Crescimento celular em meio com estresse máximo de etanol e enriquecido com espermidina	75

## Lista de Tabelas

Tabela 1: descrição das linhagens estudadas	33
Tabela 2: Agrupamento das linhagens estudadas e suas tolerâncias máximas ao EtOH	52
Tabela 3: Resumo das métricas de montagem da linhagem BMA64-1A considerando 196 montagens	53
Tabela 4: Número de Genes Diferencialmente expressos por linhagem	56
Tabela 5: Genes diferencialmente expressos comumente relacionados a alguns estressores	60
Tabela 6: Processos biológicos obtidos pelo REVIGO em grupos HT e LT	62
Tabela 7: KEGG pathways significativas detectadas pelo método GAGE	63
Tabela 8: Comparação dos dados de ensaios ROS, H2AX e SDH	68
Tabela 9: Ensaio de diferença na densidade óptica	69
Tabela 10: Porcentagem de células acumulando danos ao DNA (ensaio utilizando o anticorpo H2AX)	70
Tabela 11: Diferença entre os quadrantes do ensaio de anexina	70
Tabela 12: Estatísticas gerais das redes específicas	71
Tabela 13: Diferença estatística em medidas topológicas dos nós	72
Tabela 14: Performance preditiva em diferentes matrizes	76

## Lista de Siglas

ADH – Aldeído desidrogenase

ALD – Álcool desidrogenase

BC - *burtconstraint*

BET - *betweenness*

C - *coreness*

CLOS - *closeness*

CO2 - Dióxido de carbono

DAM – *Differentially Abundant Metabolites* (Metabólitos diferencialmente abundantes)

ECC - *eccentricity*

EIG - *eigenvectors*

EtOH - etanol

FDR - *False discovery rate*

GO - *Gene ontology term*

GRN - *Gene regulatory network* (rede regulatória genética)

HS - *hub score*

HT - *High tolerant*

K - *degree*

Knn - *Average nearest neighbor degree*

Inc-p - rede de interação *long-non-coding RNAs* com proteínas

LT - *Low tolerant*

Matriz CT - Matriz controle/tratamento

Matriz HTLT - Matriz Altamente tolerante/pouco tolerante

NRBP - *National Bioresources Project*

PPI - *Protein-protein interaction network* (rede de interação proteína-proteína)

PR - *page rank*

ROS - *Reactive oxygen species* (espécies reativas de oxigênio)

S - *strength*

SC - *subgraph centrality*

SGD – *Saccharomyces genome database*

T – *transitivity*

TE – *transposable element* (elemento transponível)

YPD - *yeas extract peptone*

## Sumário

<b>1. Introdução</b>	<b>18</b>
1.1. Biocombustíveis e o fenômeno de tolerância ao etanol	18
1.2. Tolerância ao etanol em <i>Saccharomyces cerevisiae</i>	21
1.3. Biologia de sistemas, transcriptomas e seus empregos no estudo de tolerância ao etanol	25
1.4. Hipóteses	30
<b>2. Objetivos</b>	<b>31</b>
2.1. Objetivos geral	31
2.2. Objetivos específicos	31
<b>3. Materiais e métodos</b>	<b>32</b>
3.1. Experimentos de tolerância máxima ao etanol	32
3.2. Extração de biomoléculas, controle de qualidade e quantificação	34
3.3. Ensaio de biologia celular	36
3.3.1. Quantificação de células com danos ao DNA e de células com acumulação de espécies reativas de oxigênio	37
3.3.2. Ensaio de Succinato Desidrogenase	38
3.3.3. Viabilidade Celular	39
3.4. Sequenciamento, montagem e anotação do genoma	40
3.5. Análise de RNA-Seq e metabolomas: expressão diferencial , Enriquecimento de GOs e anotações KEGG	43
3.6. Tratamento com espermidina	45
3.7. Rede global integrada e análises estatísticas	47
3.8. Machine Learning aplicado a redes biológicas	48
<b>4. Resultados</b>	<b>52</b>
4.1. O nível de tolerância ao etanol	52
4.2. Montagem e anotação genômica	52
4.3. Análise de abundância diferencial de metabólitos	53
4.4. Análise de expressão diferencial	56
4.5. Citometria de fluxo e ensaio SDH	67
4.6. Rede global integrada e análises estatísticas	71
4.7. Experimentos de crescimento celular acrescidos de espermidina	74
4.8. Rede global integrada e análises de Machine Learning	76
<b>5. Discussão</b>	<b>77</b>
5.1. Remodelando o <i>diauxic shift</i> : mecanismo fundamental para a tolerância ao etanol	77
5.2. A espermidina tem um papel importante para a tolerância ao etanol	80
5.3. O estresse por etanol ocasiona um expressivo rewiring das redes: os módulos basais tendem a ser os que se mantêm mais ativos	82
5.4. Rede global integrada e análises de Machine Learning	83
5.5. Visão geral da tolerância ao etanol em <i>S. cerevisiae</i>	84
<b>6. Conclusões</b>	<b>86</b>
<b>7. Referências</b>	<b>87</b>

## **Apêndices**

1	Tabela completa com Termos do REVIGO	<b>99</b>
2	Genes Diferencialmente Expressos e Metabólitos Diferencialmente Abundantes	<b>102</b>
3	Estudo Piloto de Machine Learning aplicado a redes Biológicas	<b>103</b>
4	Projetos realizados durante o doutorado	<b>113</b>

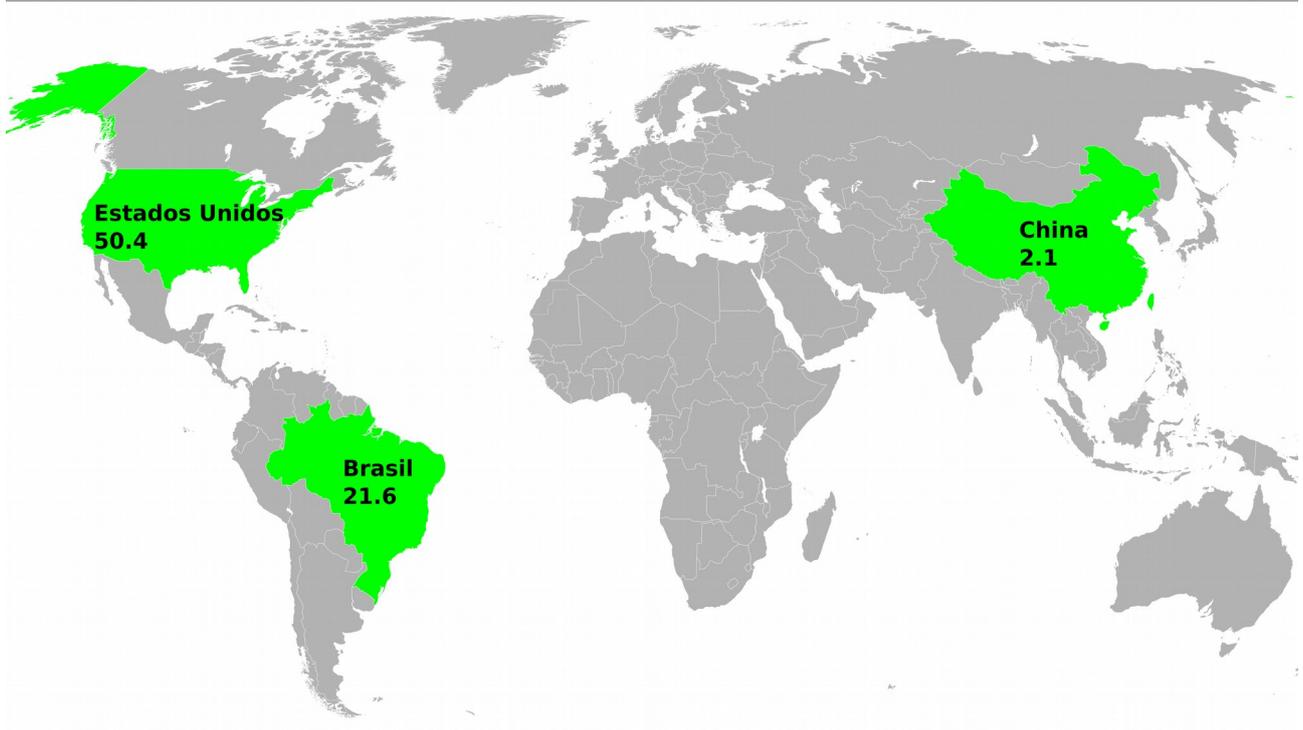
## 1. Introdução

### 1.1. Biocombustíveis e o fenômeno de tolerância ao etanol

O crescimento das demandas energéticas acompanham o crescimento da população mundial, e atualmente as fontes primárias de energia se baseiam em combustíveis fósseis não renováveis (GUPTA; VERMA, 2015). Os combustíveis fósseis são a fonte de aproximadamente 80% da energia consumida no mundo, conseqüentemente, o preço e a crescente demanda destes combustíveis afeta diretamente a economia global (CHAKRABORTY et al., 2012). Este fato desperta o interesse no desenvolvimento do mercado de biocombustíveis em vários países, com o objetivo de amenizar os efeitos econômicos causados pelo preço dos combustíveis fósseis, mitigar as emissões de CO<sub>2</sub> e atingir as metas estabelecidas em acordos internacionais sobre o meio ambiente (MUSSATTO et al., 2010).

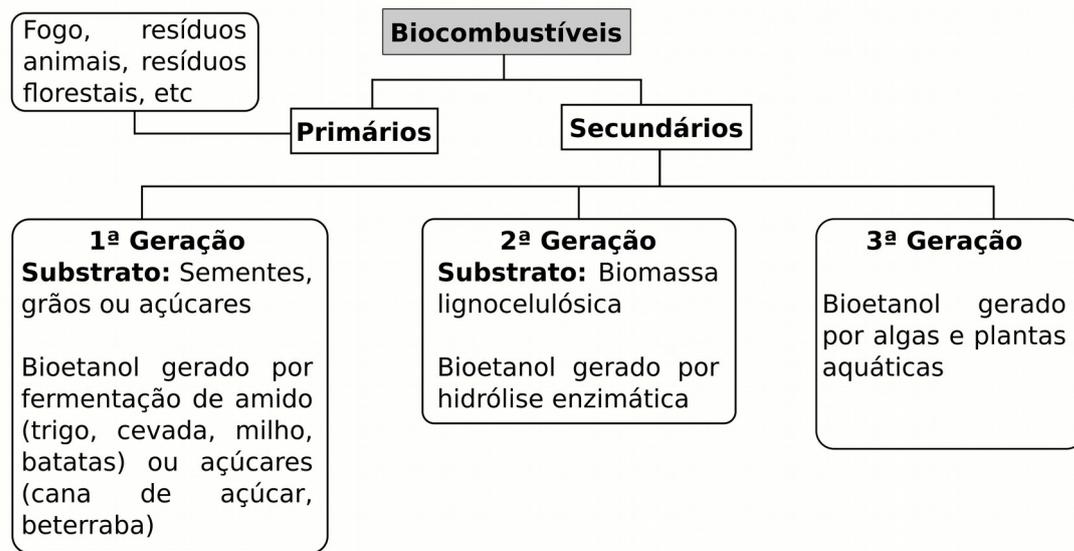
Os biocombustíveis constituem uma forma sustentável de obtenção de energia e podem ser produzidos a partir de fontes agrárias e biomassa de plantas, como o trigo, beterraba e milho; e de alguns resíduos biológicos, como palha, madeira e mesmo alguns resíduos domésticos (DEMIRBAS, 2017). Dentre os biocombustíveis, o bioetanol está entre os mais importantes por ser caracterizado como um dos mais promissores substitutos da gasolina no mercado de combustíveis para transporte (CHAKRABORTY et al., 2012; GUPTA; VERMA, 2015).

A importância do bioetanol se torna mais clara ao observar que os Estados Unidos da América e o Brasil são os dois maiores produtores de bioetanol e estima-se que serão os dois principais consumidores deste biocombustível em 2024 (GUPTA; VERMA, 2015; DEMIRBAS, 2017) (**Figura 1**).



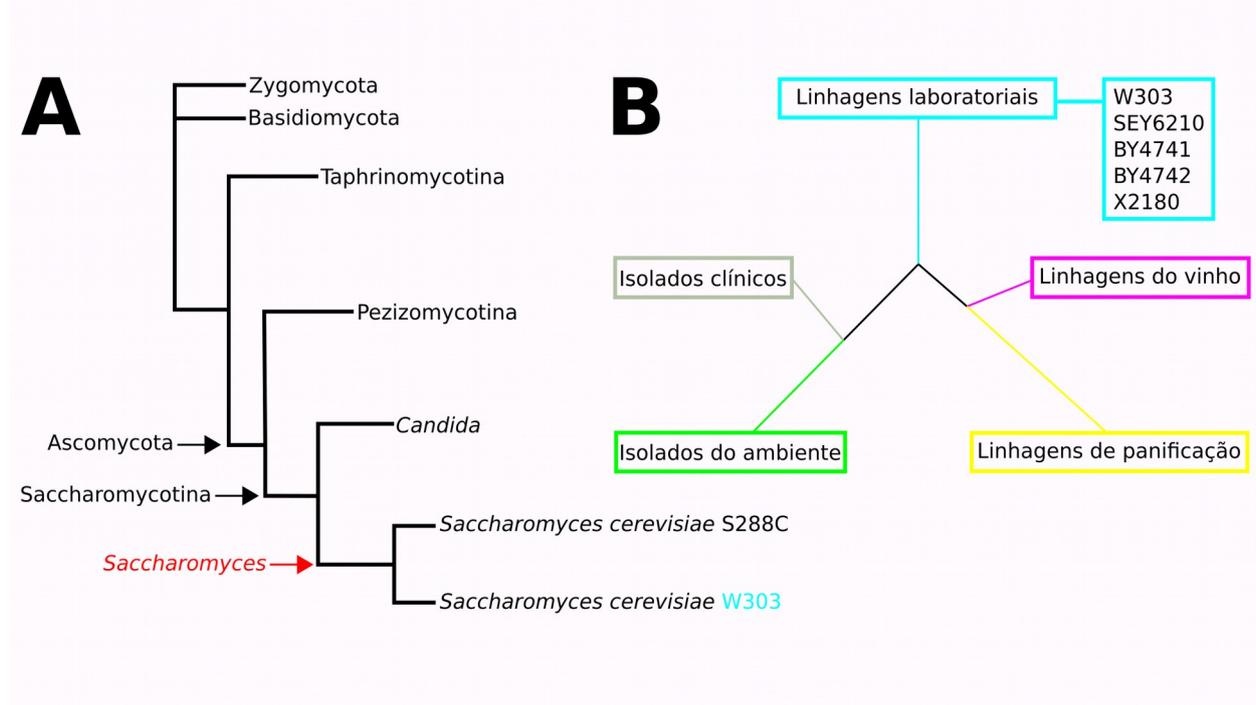
**Figura 1:** Produção de bioetanol (bilhões de litros) em 2012. Fonte (GUPTA; VERMA, 2015).

O bioetanol é um biocombustível secundário obtido principalmente pelas tecnologias de primeira geração, a fermentação alcoólica (**Figura 2**). Neste processo, geralmente os açúcares de seis carbonos de sementes, grãos ou da cana-de-açúcar são convertidos em etanol por enzimas da levedura (NIGAM; SINGH, 2011).



**Figura 2:** Classificações do bioetanol como biocombustível. Fonte: (NIGAM; SINGH, 2011).

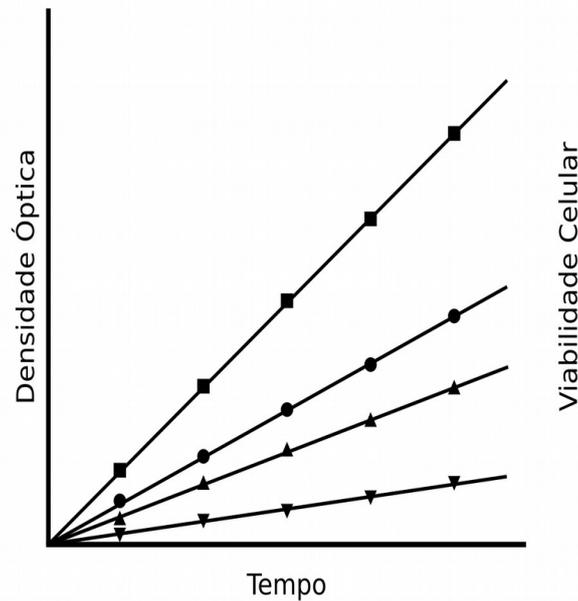
A levedura *Saccharomyces cerevisiae* possui diversas linhagens laboratoriais e depósitos em bancos de dados, pertence ao filo Ascomycota, subfilo Saccharomycotina e ao gênero *Saccharomyces* (**Figura 3A-C**). Esta levedura destaca-se como os organismos mais utilizados para processos industriais de produção de bioetanol por apresentar alta capacidade fermentativa e de crescimento em condições anaeróbias, possuir alta tolerância às condições industriais, tais como temperatura e pH, e também apresentar uma tolerância relativamente alta ao etanol e a outros inibidores (MUSSATTO et al., 2010; DEMEKE et al., 2013).



**Figura 3:** Filogenia de *Saccharomyces cerevisiae*. **A:** Localização de *S. cerevisiae* em Saccharomycetales, Vermelho: Clado de *Saccharomyces*; Azul: Linhagem laboratorial W303 de *S. cerevisiae*; **B:** Filogenia geral de linhagens do *Saccharomyces* Genome Database. Modificado de: (BUTLER, 2010; SONG et al., 2015; GALLONE et al., 2016).

## 1.2. Tolerância ao etanol em *Saccharomyces cerevisiae*

Durante a produção de bioetanol a levedura é exposta a alguns tipos de estresses destacando-se o osmótico, por aquecimento, o oxidativo e o aumento da concentração de etanol (AUESUKAREE, 2017). Apesar de *S. cerevisiae* ser comumente utilizada na produção de bioetanol, a alta concentração deste composto compromete a sobrevivência da levedura afetando sua taxa de crescimento (STANLEY et al., 2010a) (**Figura 4**).



**Figura 4:** Efeito de várias concentrações de etanol no crescimento de *Saccharomyces cerevisiae*. As linhas de tendência exemplificam o comportamento das células quando inoculadas em meio definido apenas (■) ou em meio definido contendo 5% v/v (●), 7% v/v (▲) ou 10% v/v (▼) de etanol adicionado. Fonte: (CHANDLER et al., 2004).

Assim, o acúmulo de etanol constitui um dos estressores no processo de fermentação. Dentro das células, os efeitos comuns da toxicidade por etanol incluem a inibição da divisão celular, diminuição do seu volume, aumento da taxa de morte e redução da vitalidade das células (JIA; ZHANG; LI, 2010; STANLEY et al., 2010a, 2010b), afetando negativamente a produção em escala industrial. Conseqüentemente, para um maior custo-benefício na produção de bioetanol, são necessárias linhagens com maior tolerância à este composto, ou seja, que tenham um potencial de sobrevivência maior durante uma exposição crônica a esta substância (JIA; ZHANG; LI, 2010; STANLEY et al., 2010a; YU et al., 2012).

A exposição de *S. cerevisiae* ao estresse por etanol afeta primeiramente a integridade da membrana plasmática, alterando sua fluidez e permeabilidade às espécies iônicas levando a uma perda do potencial eletroquímico transmembrana e, em seguida,

acidificando a célula e seus vacúolos (CHANDLER et al., 2004; DING et al., 2009; MA; LIU, 2010; NAVARRO-TAPIA et al., 2016); desse modo, a homeostase intracelular encontra-se seriamente comprometida. Altas concentrações de etanol também podem causar disfunção e desnaturação de proteínas, afetar a absorção de glicose, maltose, amônia e aminoácidos e, finalmente, provocar o escape de nucleotídeos, aminoácidos e potássio das células (DING et al., 2009; MA; LIU, 2010). Análises de *microarray* confirmam que muitos destes sistemas e mecanismos sofrem ajustes em seus perfis de expressão frente ao estresse por etanol (ALEXANDRE et al., 2001; CHANDLER et al., 2004; STANLEY et al., 2010b).

Mais de 200 descrições ontológicas já foram identificadas como participantes na resposta ao estresse ou tolerância ao etanol, incluindo: 1- a biossíntese de moléculas tais como prolina, triptofano e proteínas; 2- o metabolismo de aminoácidos, nucleotídeos, lipídios, ergosterol e ácidos graxos; 3- e os sistemas de transporte, crescimento e ciclo celular (DING et al., 2009; MA; LIU, 2010; STANLEY et al., 2010b; NAVARRO-TAPIA et al., 2016).

Levando em consideração o interesse científico e econômico sobre a característica de tolerância ao etanol (DING et al., 2009), diversos trabalhos de engenharia genética buscam o desenvolvimento de linhagens de *S. cerevisiae* mais resistentes a este estressor (ALPER et al., 2006; TEIXEIRA et al., 2009; LEWIS et al., 2010; MUSSATTO et al., 2010; SWINNEN et al., 2012). Para isso, normalmente são utilizadas estratégias racionais ou randômicas que demandam de tempo e investimentos, tais como o *screening* de fenótipos, mutações e *knockout* aleatórios (JIA; ZHANG; LI, 2010).

Estudos como os acima citados, permitiram reportar que os genes relacionados à tolerância ao etanol normalmente estão associados a mais de uma função (MA; LIU,



*cerevisiae*. Proteínas codificadas por genes significativamente regulados positivamente estão localizadas na parede celular, membrana, núcleo, mitocôndria e citoplasma. Proteínas de choque térmico (HSPs) são principalmente descritas como chaperonas protegendo e mantendo funções de proteínas em múltiplos locais. A função dos produtos gênicos é classificada com base na ontologia gênica. Fonte: (MA; LIU, 2010).

Os fatos supracitados deixam claro que a relação entre os elementos moleculares-chave que sustentam a aclimatação da levedura ao etanol e àqueles associados a determinar sua tolerância, são ainda pouco esclarecidos (STANLEY et al., 2010b; SWINNEN et al., 2012). Isso provavelmente está relacionado ao fato de os estudos tradicionais avaliarem as funções gênicas abordando poucos genes por vez, desconsiderando que os produtos gênicos atuam como componentes de uma rede integrada (GE; WALHOUT; VIDAL, 2003). O mecanismo de tolerância ao etanol está associado a muitos genes e possui múltiplas camadas de funções e interações complexas (MA; LIU, 2010) (**Figura 5**).

### 1.3. Biologia de sistemas, transcriptomas e seus empregos no estudo de tolerância ao etanol

Tradicionalmente os estudos de biologia molecular utilizam abordagens reducionistas, as quais objetivam elucidar os princípios biológicos a partir de uma ou poucas moléculas (como genes, proteínas e metabólitos) por vez (GE; WALHOUT; VIDAL, 2003). No entanto, o estudo de moléculas isoladas não permite a compreensão de características multifatoriais, sendo necessário uma abordagem sistêmica de análises (BRUGGEMAN; WESTERHOFF, 2007).

Assim, para aprimorar o conhecimento sobre a função dos módulos que regem os sistemas biológicos de maneira integrada, é imprescindível combinar múltiplos conjuntos de dados (GE; WALHOUT; VIDAL, 2003), comparar os resultados obtidos com outros já existentes e verificar se os resultados se repetem em conjuntos de dados auxiliares, atingindo assim um “padrão ouro” para estas análises (RITCHIE et al., 2015). Por

preencher os requisitos supracitados, a metodologia de análise de redes vem sendo cada vez mais utilizada. Nesta abordagem os vértices (também chamados de nós) geralmente constituem os componentes biológicos e as arestas (também chamadas de conexões) representam as suas interações.

Análises de redes são amplamente utilizadas para descrever diversos sistemas (LUSCOMBE et al., 2004). Dentre suas vantagens, elas independem do tipo de dados e permitem sua aplicação em análises funcionais integradas (PIRUZIAN et al., 2010). Além disso, testes estatísticos sobre essas redes permitem que os resultados geralmente sejam interpretados e comparados, assim os mecanismos capazes de perturbá-los podem ser identificados (CREIXELL et al., 2015).

Trabalhos utilizando metodologias de análise de redes em *S. cerevisiae* foram aplicados em estudos de redes regulatórias (ver (LEE et al., 2002; LUSCOMBE et al., 2004; HERRGARD, 2006)), redes metabólicas (ver (FAMILI et al., 2003)), redes globais (ver (SAID et al., 2004; REGULY et al., 2006; YU et al., 2008)), redes do ciclo celular (LI et al., 2004) e em apenas dois trabalhos abordou-se a tolerância ao etanol (KASAVI et al., 2014, 2016).

Os estudos previamente citados (KASAVI et al., 2014, 2016) exploram as propriedades das redes de maneira estática. No entanto, existe também a necessidade de se compreender a conectividade das redes de maneira dinâmica (LEE et al., 2012). Como exemplo, já foi observado que diferentes tratamentos podem, sinergicamente, alterar o estado intrínseco das células por meio de reconexões dinâmicas nas redes regulatórias (LUSCOMBE et al., 2004; LEE et al., 2012). Desta forma, não é aconselhável desconsiderar que as interações em uma rede biológica variem de acordo com os diferentes estímulos (LUSCOMBE et al., 2004; CREIXELL et al., 2015), o que pode alterar suas medidas topológicas.

Neste contexto, somente uma análise de redes dinâmicas (considerando aqui a dinâmica entre diferentes condições) foi efetuada em *S. cerevisiae* (LUSCOMBE et al., 2004). Neste trabalho os autores verificaram que as respostas ambientais em medições topológicas globais bem conhecidas em redes, tais como os *degrees* de um vértice, indicam alterações dinâmicas transitórias nas interações entre as proteínas nas diferentes condições avaliadas. Estes conceitos de alterações topológicas e estados de transição podem ser aplicados a outras redes biológicas, incluindo sub-sistemas complexos (LUSCOMBE et al., 2004) como é o caso da característica de tolerância ao etanol.

No contexto do uso de análises em *high-throughput* de RNAs visando a tolerância ao etanol, há relatos de estudos realizados com as bactérias *Zymomonas mobilis* ZM4 (YANG et al., 2013), *Clostridium thermocellum* (YANG et al., 2012) e cianobactéria *Synechocystis* sp. (WANG et al., 2012). Dos poucos trabalhos envolvendo transcriptomas com foco na tolerância ao etanol em *S. cerevisiae*, a metodologia empregada foi a de *microarray* (ALEXANDRE et al., 2001; CHANDLER et al., 2004; STANLEY et al., 2010b) e nenhum deles realizou análises sistêmicas a fim de avaliar eventuais diferenças entre o controle e o tratamento com etanol (o mesmo vale para os trabalhos com bactérias e cianobactérias supracitados). Adicionalmente a tecnologia de *microarray* é conhecida por requerer uma alta quantidade de RNA, gerar mais ruídos que o RNA-Seq e não permite o estudo de isoformas e de diferentes alelos (WANG; GERSTEIN; SNYDER, 2009). Por outro lado, análises de RNA-Seq evitam alguns dos problemas citados e apresentam um alcance dinâmico muito maior no que diz respeito aos níveis de expressão (FU et al., 2009; WANG; GERSTEIN; SNYDER, 2009; XU et al., 2013).

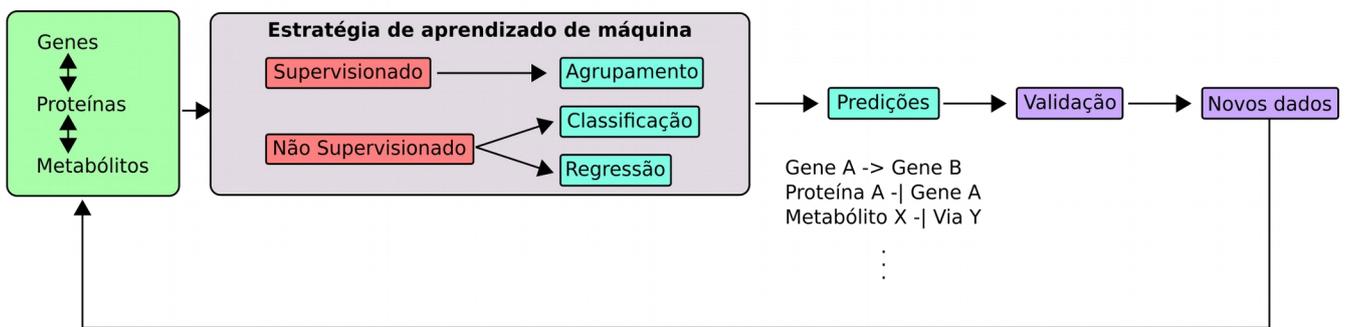
No que se refere a redes, Kasavi *et al.* (KASAVI et al., 2014, 2016) utilizando redes de interações entre proteínas (PPIs), redes regulatórias e dados de *microarray*, identificaram 17 novos genes candidatos relacionados a tolerância ao etanol. Análises de

*microarray* foram efetuadas para verificar a influência da deleção de dois destes genes. No entanto, as linhagens avaliadas possuíam apenas 1% de divergência quanto à tolerância ao etanol, não houve tratamento com etanol antecedendo a extração de RNA no primeiro experimento e, além disso, as três linhagens testadas tinham o mesmo *background* genético da linhagem BY4743. Logo, apesar de obterem resultados interessantes, não houve nenhuma tentativa de compreender a importância do comportamento dinâmico das redes geradas junto aos perfis de expressão obtidos.

Assim, ao observar os estudos anteriores, é notável a abundância de informação disponível a cerca do fenômeno de tolerância ao etanol. Infelizmente, devido a quantidade e complexidade, a análise integrada destes dados torna-se inviável por meio de métodos convencionais. Neste contexto, ferramentas e técnicas são necessárias para lidar com as grandes quantidades de informações biológicas sendo geradas e o *Machine learning* tem se destacado nesse campo por permitir a análise destes conjuntos de dados complexos (CAMACHO et al., 2018).

O *Machine learning* é uma área derivada da inteligência artificial, no qual um programa (algoritmo) aprende a respeito de uma tarefa e otimiza seu aprendizado através de uma medida de performance (DAS et al., 2015) (**Figura 6**). Diferentes algoritmos podem ser usados dependendo do conjunto de dados e da tarefa a ser executada. No caso de uma tarefa de predição na qual previamente sabe-se os rótulos/classes (*labels*) dos dados, pode-se trabalhar com o aprendizado supervisionado (*supervised learning*) para criar modelos de classificação ou regressão. Caso o objetivo seja encontrar estrutura nos dados sem nenhuma informação prévia, pode-se utilizar o aprendizado não-supervisionado (*unsupervised learning*) para executar uma tarefa de agrupamento (*clustering*) (DAS et al., 2015; CAMACHO et al., 2018). Diversos algoritmos de *Machine learning* estão disponíveis, sendo que alguns deles permitem de forma

simples a interpretação dos modelos gerados, tal como observa-se no uso das árvores de decisão. Por outro lado, o uso de redes neurais artificiais geram dados de difícil interpretação (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2006). Finalmente, apesar de técnicas de *Machine learning* estarem sendo aplicadas a diversos tipos de dados biológicos (DAS et al., 2015; CAMACHO et al., 2018), não há nenhum estudo onde estas técnicas tenham sido aplicadas à tolerância ao etanol em *S. cerevisiae*.



**Figura 6:** Tipos de dados e técnicas de *Machine learning*. De acordo com os objetivos e os dados fornecidos uma tarefa preditiva (*Supervised learning*) ou uma tarefa de agrupamento (*Unsupervised learning*) pode ser utilizada. Fonte: (CAMACHO et al., 2018).

Deste modo, essa tese objetiva identificar assinaturas sistêmicas associadas à tolerância ao etanol por meio do uso de transcriptoma e metaboloma, técnicas de bioinformática, inteligência artificial, biologia celular e redes dinâmicas, ou seja, análises de alterações dinâmicas nas propriedades dos nós e das conexões das redes em diferentes condições (presença ou ausência de etanol). Experimentalmente, diferentes linhagens de *S. cerevisiae* com diferentes tolerâncias ao etanol foram submetidas à esse estressor. Os transcriptomas e metabolomas dessas linhagens foram obtidos durante o estresse máximo e tais dados foram explorados também em nível sistêmico, integrando dados de diferentes fontes.

#### 1.4. Hipóteses

- Linhagens altamente tolerantes e pouco tolerantes ao etanol devem exibir uma sobreposição parcial dos sistemas envolvidos na tolerância. A análise desses sistemas utilizando algoritmos de inteligência artificial sobre as características topológicas dos genes dentro das redes de interação poderá distinguir esses padrões para a classificação das linhagens.
- A análise de padrões, utilizando algoritmos de inteligência artificial, integrada a análise de redes, genômica e transcriptômica, permitirão extrair informações que indiquem quais sistemas e quais alterações são as mais relevantes para a característica escolhida.

## 2. Objetivos

### 2.1. Objetivos geral

Identificar assinaturas sistêmicas associadas à tolerância ao etanol em linhagens de *Saccharomyces cerevisiae*.

### 2.2. Objetivos específicos

1. Obter dados de transcriptomas e metabolomas para diferentes linhagens de *S. cerevisiae* com diferentes níveis de tolerância máxima ao etanol;
2. Identificar quais são os genes/metabólitos comuns, em termos de nível de expressão, para cada grupo de linhagens a partir dos transcriptomas e metabolomas;
3. Identificar genes homólogos dentro de cada grupo e que sejam diferencialmente expressos entre os grupos muito e pouco tolerantes ao etanol;
4. Reconstruir uma rede integrada de *S. cerevisiae* a partir de redes já disponíveis em banco de dados;
5. Utilizar os transcriptomas e a rede integrada para criar as redes de cada linhagem no estado de controle (ausência de etanol no meio) e tratamento (presença de etanol no meio);
6. Analisar e comparar os diferentes estados das redes;
7. Atribuir análises de enriquecimento de termos GO;
8. Integrar dados de transcriptoma, metaboloma e de biologia celular para avaliar a tolerância ao etanol;
9. Criar um modelo para mostrar as características mais importantes para a tolerância;
10. Buscar padrões nos dados gerados que expliquem a tolerância ao etanol.

### 3. Materiais e métodos

#### 3.1. Experimentos de tolerância máxima ao etanol

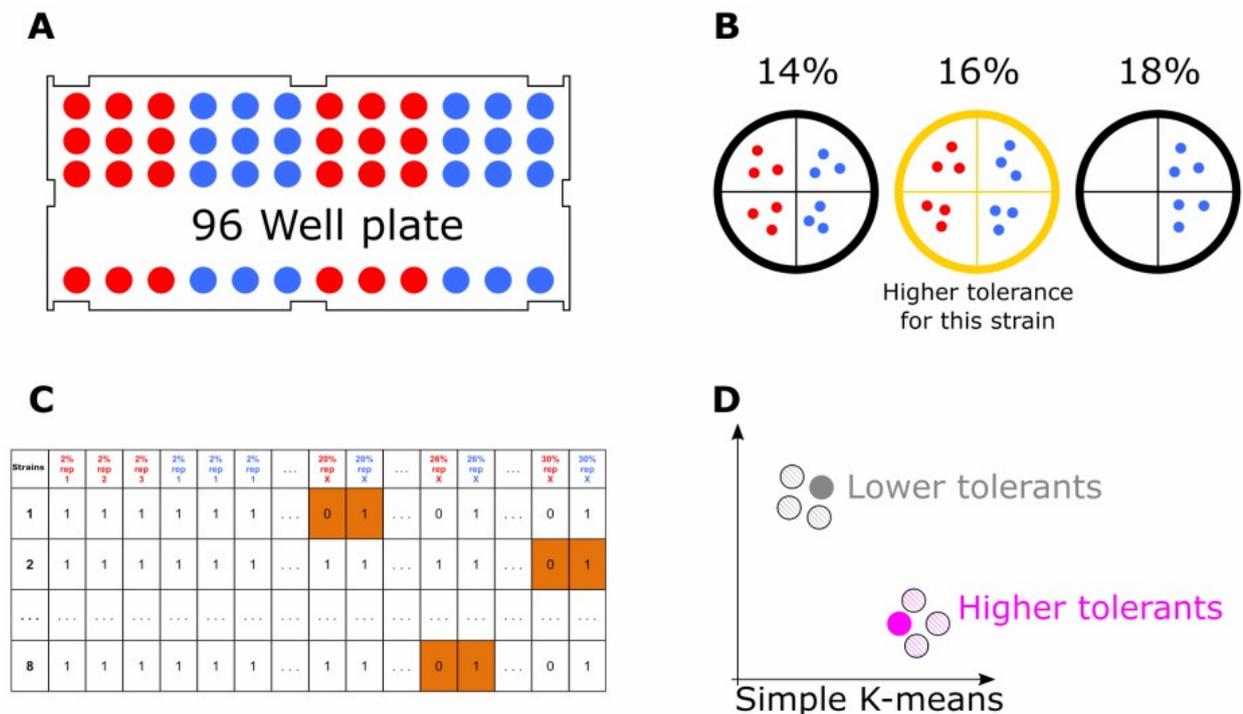
O protocolo de caracterização de tolerância máxima ao etanol para as linhagens aqui estudadas foi realizado de acordo com Lewis et al. (LEWIS et al., 2010) por ALMEIDA (ALMEIDA, 2017). O maior nível de estresse por etanol (EtOH) suportado pelas células foi definido para 13 diferentes linhagens de *S. cerevisiae* obtidas da Euroscarf (*European Saccharomyces cerevisiae archive for Functional Analysis*) e NRBP (*National Bioresources Project*). Para isso as linhagens em fase exponencial de crescimento com OD<sub>600nm</sub> de 0.2 foram colocadas em placas de 96 poços, centrifugadas e o sobrenadante descartado. Meio YPD com diferentes concentrações (de 2% à 32%, v/v) de EtOH (tratamento), bem como com solução fisiológica (controle), foram adicionados nos poços e as células foram tratadas por 1 h à 30 °C (**Figura 7A**). Após os tratamentos, o conteúdo de cada poço foi replicado em placas com YPD e crescidas por 24 h à 30 °C. As porcentagens nas quais não se observa crescimento celular após o tratamento mas sim no controle, reportam o nível de estresse que tornou inviável o crescimento. Contudo, para definir o nível máximo de tolerância ao etanol, o nível de estresse anterior ao que inviabilizou o crescimento foi definido como sendo o nível máximo de tolerância ao etanol suportada para uma dada linhagem (ALMEIDA, 2017).

Posteriormente foi realizado a definição de linhagens como altamente tolerante ao EtOH (HTs) ou pouco tolerante ao EtOH (LTs) (**Figura 7B**). Para isso a matriz com as medidas de tolerância ao EtOH para 8 linhagens haplóides foi submetida para um aprendizado não supervisionado com o algoritmo *Simple K-means* (K=2) implementado no programa Weka (HALL et al., 2009) para classificar as linhagens em HT ou LT (**Figura 7C e D**). Um total de 3 linhagens determinadas como LTs (S288C, SEY6210 e BY4741) e

3 como HTs (BMA64-1A, X2180-1A e BY4742) foram utilizadas no presente estudo (Tabela 1).

**Tabela 1:** Descrição das linhagens estudadas. <sup>1</sup>: linhagem com o genoma sequenciado; **Pub. Prot:** Número de acesso do proteoma em bancos de dados públicos; **Num. Ac. Comp.:** Número de acesso da instituição da qual a linhagem foi obtida. Adaptada de ALMEIDA (ALMEIDA, 2017).

Linhagem	Pub. Prot	Num. Ac. Comp.	Ploidia(MAT locus)
BMA64-1A <sup>1</sup>	Not available	Euroscarf/20000A	a
BY4741	SGD/BY4741	Euroscarf/Y00000	a
BY4742	SGD/BY4742	Euroscarf/Y10000	alpha
S288c	SGD/S288c	NBRP/BY20118	alpha
SEY6210	SGD/SEY6210	NBRP/BY3553	alpha
X2180-1A	SGD/X2180-1 <sup>a</sup>	NBRP/BY21559	a



**Figura 7:** Desenho experimental. **A:** Determinação da tolerância ao etanol para 13 linhagens e sua classificação de acordo com suas tolerâncias em concentrações crescentes de etanol; **B:** Cultivo após 1h de tratamento; **C:** Matriz com todas as medidas; **D:** Resultado do aprendizado não supervisionado; Círculos Vermelhos e azuis: as condições de tratamento e controle, respectivamente; Caixas em laranja: maior tolerância para cada linhagem; Círculos cinza: linhagens LT; Círculos Rosa: linhagens HT. Imagem adaptada de ALMEIDA (ALMEIDA, 2017).

Posteriormente, os experimentos sob estresse máximo para as 6 linhagens haplóides selecionadas (Tabelas 1 e 2) foram repetidos e as amostras foram congelados à -80 °C para extração de DNA e RNA. Para a extração de metabólitos, as células depois

do tratamento de estresse máximo foram imediatamente submetidas ao protocolo de extração de metabólitos (descrito mais adiante).

### 3.2. Extração de biomoléculas, controle de qualidade e quantificação

Para a extração total de RNAs, 1,5mL de cada cultura celular congelada foi lavada com solução fisiológica e a extração de RNA foi efetuada usando o protocolo Trizol® padrão, com exceção do uso da enzima liticase (para degradação da parede celular). A qualidade dos RNAs foi verificada usando gel de agarose 1% em tampão TBE e Bioanalyzer. As quantidades de RNAs foram estimadas usando Nanodrop e Qubit. As amostras foram tratadas com o kit TURBO DNase (ThermoFisher®) e armazenados à -80 °C.

A linhagem BMA64-1A, a HT mais tolerante, não possuía o genoma sequenciado. Então, o DNA foi extraído usando Fenol-Clorofórmio e a qualidade e quantidade foram estimadas com os mesmos equipamentos utilizados para as amostras de RNA, para então serem enviadas para o sequenciamento genômico pela GenOne (Rio de Janeiro, Brasil) (detalhes descritos posteriormente).

Adicionalmente foi efetuada a extração total de metabólitos seguido de espectrometria de massa. Para essa extração, 120 mL de células tratadas e controle foram imediatamente transferidas para 5 tubos Falcon de 50 mL e o volume foi preenchido para 50 mL com solução fisiológica. Os tubos foram centrifugados (3.000 RPM), o sobrenadante foi descartado e 10 mL de solução fisiológica foram adicionados e homogeneizados. Depois, 40 mL de metanol frio (-40 °C da solução de extinção) foram adicionados e homogeneizados por inversão 3 vezes e repousaram em inibidor por 5 min. As amostras foram centrifugadas (3.000 RPM e -40 °C), em seguida, o sobrenadante foi descartado, e ~0,2g de *pellet* foi transferido para microtubos de 1,5 mL. Depois, o *pellet*

foi lavado adicionando 1 mL de MeOH:H<sub>2</sub>O (50:50 v/v) e repousou durante 4 min em gelo, misturando durante 1 min e finalmente repousando em -80 °C durante 30 min; o passo anterior foi repetido 3 vezes. Os tubos foram centrifugados (15.000 g durante 15 min a 4 °C) e o fluído foi filtrado utilizando uma membrana de PVDF (0,22 µm) utilizando uma seringa. As amostras foram transferidas para frascos e posteriormente secas. Todo esse protocolo foi adaptado de Kim *et al.* (KIM *et al.*, 2013) e o foi realizado para as 6 linhagens selecionadas (**Tabela 1**). Pelo menos 5 réplicas biológicas foram coletadas e réplicas técnicas também foram realizadas para algumas amostras escolhidas aleatoriamente.

As amostras foram derivatizadas, onde os compostos tem sua estrutura alterada para permitir sua detecção, de acordo com Hoffman *et al.* (HOFFMAN *et al.*, 2010) com 30 µL de cloridrato de metoxiamina (15 mg mL<sup>-1</sup>) em piridina durante 16 h em temperatura ambiente. As amostras foram trimetilsililadas por adição de 30 µL de N-metil-N-(trimetilsilil) trifluoroacetamida (MSTFA) contendo trimetilclorossilano a 1% (TMCS) e a mistura resultante repousou temperatura ambiente durante 1 h. Após a sililação, adicionou-se 30 µL de heptano. Compostos de referência isotópicos estáveis [1 mg mL<sup>-1</sup> de cada um de (<sup>13</sup>C<sub>3</sub>) - ácido mirístico, (<sup>13</sup>C<sub>4</sub>) - ácido palmítico e (<sup>2</sup>H<sub>4</sub>) - ácido succínico] foram adicionados nas amostras antes da derivatização e usados como padrão externo para controle de qualidade. As amostras derivatizadas foram analisadas de acordo com Gullberg *et al.* (GULLBERG *et al.*, 2004). Amostras de controle em branco e uma série de n-alcenos (C<sub>12</sub> – C<sub>40</sub>), que permitiram calcular os índices de retenção (SCHAUER *et al.*, 2005), também foram utilizados.

Um microlitro de cada amostra derivatizada foi injetado sem divisão em um cromatógrafo a gás 7890A (Agilent Technologies, Santa Clara, EUA) acoplado a um amostrador automático Comb-xt (Leap Technologies, Carrboro, EUA). A temperatura do injetor foi de 280 °C, a taxa de fluxo de purga do septo foi de 20 mL min<sup>-1</sup> e a purga foi

ligada após 60 s. A taxa de fluxo de hélio gasoso através da coluna foi de  $1 \text{ mL min}^{-1}$ , a temperatura da coluna foi mantida a  $80 \text{ }^\circ\text{C}$  durante 2 min, depois aumentada de  $15 \text{ }^\circ\text{C min}^{-1}$  até  $305 \text{ }^\circ\text{C}$  e mantida nesta temperatura durante 10 min. O efluente da coluna foi introduzido na fonte de íons de um GC×GC/TOFMS (Pegasus 4D, Leco Corp., St. Joseph, EUA) equipado com duas colunas de sílica fundida, uma primeira dimensão (Agilent DB-5) com comprimento de 20 m ( $0,18 \text{ mm}$  de diâmetro interno x  $0,18 \text{ }\mu\text{m}$  de filme) e a segunda coluna (RXT-17) de tamanho 0,96 m ( $0,10 \text{ mm}$  de diâmetro interno x  $0,10 \text{ }\mu\text{m}$  de filme). A linha de transferência e as temperaturas da fonte de íons foram  $280$  e  $250 \text{ }^\circ\text{C}$ , respectivamente. Íons foram gerados por um feixe de elétrons de  $70 \text{ eV}$  a uma corrente de ionização de  $2,0 \text{ mA}$ , e 10 espectros  $\text{s}^{-1}$  foram registrados na faixa de massa  $m/z$  45-800.

### 3.3. Ensaio de biologia celular

Nessa etapa, a quantificação de células com acumulação significativa de espécies reativas de oxigênio, danos no DNA e análises de viabilidade celular foram conduzidas por meio de citometria de fluxo. Para isso, usou-se kits comerciais para as marcações das células e as quantificações foram realizadas por meio do citômetro de fluxo BD FACSCalibur:4-color. Ainda, também foi quantificada a atividade da enzima succinato desidrogenase por um método colorimétrico. Todas as análises foram realizadas tanto para as amostras controle como para as amostras tratamento (etanol na concentração máxima suportada pela linhagem).

### 3.3.1. Quantificação de células com danos ao DNA e de células com acumulação de espécies reativas de oxigênio

A porcentagem de células com danos significativos no DNA foi medida utilizando ensaio de citometria de fluxo com H2AX. Para isso, as amostras tratadas e controles foram transferidas para tubos de 15 mL, centrifugadas a 4 °C durante 3 minutos a 1.600 x g e o sobrenadante descartado. O sedimento de cada amostra foi ressuspensão em 10 mL de etanol a 70%, incubado durante 10 minutos para fixação, depois centrifugado durante 3 minutos a 1.600 x g e o sobrenadante foi descartado. As amostras foram lavadas 2 vezes em 10 mL de citrato de sódio 50 mM, centrifugadas durante 3 minutos a 1.600 x g e o sobrenadante foi descartando. Em seguida, 1 mL de citrato de sódio 50 mM foi adicionado, o sedimento foi ressuspensão e transferido para 1,5 mL de tubo cônico. As amostras foram centrifugadas 3 minutos a 1.600 x g e o sobrenadante descartado. Em cada amostra foram adicionados 100 µL de tampão de lise (50 mM Tris, 10 mM EDTA, 1 M sorbitol e 1 U/µL de Liticase), homogeneizados e incubados por 10 min a 37 °C. Após a incubação, 500 µL de 1X PBS foram adicionados, centrifugados a 2.500 x g por 3 minutos e o sobrenadante descartado. As leveduras foram permeabilizadas pela adição de 100 µL de BD-Cytofix/Cytoperm-Plus-Permeabilization-Buffer (BD Pharmingen, EUA) e incubadas a 37 °C por 30 minutos. Subsequentemente, foram adicionados 500 µL de 1X PBS, homogeneizados e centrifugados a 2.500 x g durante 3 minutos, seguido de uma nova lavagem com 1 mL de 1X PBS. O pellet foi então ressuspensão em 300 µL de 1X PBS, um volume de 100 µL foram separados para análise de autofluorescência. Em seguida, 5 µL de Alexa Fluor® 647 Mouse Anti-H2AX (BD Pharmingen, EUA) para verificar danos ao DNA, ou 1 µL de CellROX™ Green Reagent (Thermo, USA) para verificar espécies reativas de oxigênio (ROS), foram adicionados aos 200 µL restantes e incubados por 30 minutos a 37 °C. As amostras foram centrifugadas a 2.500 x g por 3 minutos e o

sobrenadante descartado. O sedimento foi ressuspensão e lavado em 500  $\mu\text{L}$  de 1X PBS por 3 vezes, finalmente o sedimento foi ressuspensão em 400  $\mu\text{L}$  de 1X PBS para análises no citômetro de fluxo.

Ao final, a porcentagem de células com acumulação significativa de ROS tiveram as replicatas comparadas por meio do teste T de *Student* (pareado) comparando os resultados dentro de cada grupo (HT ou LT) de acordo com as situações experimentais (tratamento vs controle) ou entre grupos (HT tratamento vs LT tratamento e HT controle vs LT controle).

### 3.3.2. Ensaio de Succinato Desidrogenase

As linhagens em situação de controle e tratamento tiveram a atividade das enzimas succinato desidrogenase quantificadas. Para isso, as mostras tratadas e controles foram transferidas para tubos de 2 mL, centrifugadas a 4 °C durante 3 minutos a 1.600 x g e o sobrenadante foi rejeitado. O sedimento foi ressuspensão e lavado duas vezes em 2 mL de 1X PBS (NaCl 137 mM, fosfato 10 mM, KCl 2,7 mM e a um pH de 7,4), sendo que na última lavagem 10 microlitros da suspensão de levedura foram separados para medir a densidade óptica para posterior normalização dos dados. Posteriormente, 75  $\mu\text{L}$  de succinato de sódio (4 M) e 75  $\mu\text{L}$  de Nitro Blue Tetrazolium (10 mg mL<sup>-1</sup>) foram adicionados a cada tratamento. Depois, todas as amostras foram simultaneamente incubadas a 37 °C sob agitação orbital durante 30 minutos. As amostras foram mantidas a ~4 °C em todas as etapas anteriores à incubação a 37 °C. Para interromper a reação de marcação, foram adicionados 20  $\mu\text{L}$  de formalina a 37% em cada amostra. Então, após 30 minutos, as amostras foram centrifugadas a 1.600 x g por 3 minutos e o sobrenadante descartado. As amostras foram lavadas duas vezes com 1 mL de 1X PBS e centrifugadas

nas mesmas condições previamente mencionadas. Subsequentemente, 75  $\mu$ L de DMSO foram adicionados e incubados a 37 °C durante 1 hora. As amostras foram centrifugadas e o sobrenadante foi recuperado para medição da absorbância em OD<sub>600nm</sub> do Nitro Blue Tetrazolium reduzido. Os dados de absorvância foram normalizados pela OD<sub>600nm</sub> inicial obtido de cada amostra.

Os resultados foram analisados utilizando o ANOVA fator único, considerando os valores das triplicadas experimentais para cada linhagem de levedura e condição experimental (tratamento vs controle).

### 3.3.3. Viabilidade Celular

Para avaliar a viabilidade celular foi utilizado o kit Annexin V FITC Apoptosis Detection (BD Pharmingen cat. Number 556547) para células tratadas e controle; as quantificações foram realizadas utilizando citometria de fluxo. Os procedimentos experimentais são reportados em ALMEIDA (ALMEIDA, 2017), porém, no presente trabalho, um maior número de replicatas biológicas foram medidas.

O teste de independência Chi-quadrado foi aplicado sobre as porcentagens de células em cada quadrante (LL, LR, UL e UR) comparando tratamento vs controle e considerando um nível de significância p-value <0.01. Para compreender a relação entre os quadrantes e os fenótipos (HT e LT, tanto no controle como no tratamento), as médias das replicatas para os quadrantes LL e UR (selecionados em virtude de possuírem as maiores variâncias comparando controle vs tratamento) foram calculadas e utilizadas em um contexto de aprendizado de máquina supervisionado por meio do Weka 3 (algoritmo de árvore de decisão J48 usando pruning) (HALL et al., 2009) atribuindo como classes dos dados as informações “HT” e “LT”, tanto “controle” como “tratamento”.

### 3.4. Sequenciamento, montagem e anotação dos genomas

Os genomas das linhagens S288C, SEY6210, BY4742, X2180-1A e BY4742 estavam disponíveis e foram obtidos em julho de 2016 do banco de dados *Saccharomyces Genome Database* (SGD) (CHERRY et al., 2012).

O genoma da linhagem BMA64-1A foi sequenciado usando o Illumina HiSeq 2500 (250 nts *paired-end*, 1 Gb de *throughput* e inserto de ~500pb). Os *reads* foram processados com o Trimmomatic v0.36 (com os parâmetros ILLUMINACLIP 00\_illumina\_adapters.fa 3:30:10, LEADING 3, TRAILING 3, SLIDINGWINDOW 4:15, MINLEN 75) e submetidos a uma montagem nos programas AbySS v2.0.2 (JACKMAN et al., 2017), IDBA v1 (PENG et al., 2010), MIRA v4.0.2, SPAdes 3.10.1 (BANKEVICH et al., 2012) e Velvet v1.2.10 (ZERBINO; BIRNEY, 2008), variando os parâmetros e estratégias de montagem.

Para montagem de cromossomos individuais, os *reads* foram mapeados com o HISAT2 v2.1.0 no genoma de referência (S288C) e os *reads* mapeados foram montados junto aos outros *reads* que mapeavam em um mesmo cromossomo. As métricas (ex. N50, número de *contigs*, genes provavelmente recuperados, tamanho total) de cada montagem foram obtidas com o QUAST v4.5 (GUREVICH et al., 2013) baseando-se no genoma da S288C como referência. Por fim, para determinar a melhor montagem foi criada uma métrica chamada de *assembly score* (eq. 1), a qual é definida pela soma das diferenças das raízes quadradas dos valores normalizados de cada métrica obtida pelo QUAST com relação ao genoma referência utilizado. O *assembly score* fornece um valor de 0 a 1 e a montagem com o *assembly score*  $\approx 0$  foi atribuída como a mais similar ao genoma de referência e foi definido como a montagem final.

$$d = \sum_{i=1}^n \sqrt{(métrica_j - métrica_i)^2} \quad (\text{eq. 1})$$

onde  $métrica_j$  é a distância euclidiana do genoma de referência e  $métrica_i$  é a distância euclidiana do genoma de comparação. O cálculo de  $d$  é realizado sobre todas as  $n$  métricas consideradas neste estudo.

A melhor montagem (melhor *assembly score*) para a linhagem BMA64-1A foi realizada separando os *reads* pareados por meio do mapeamento em cada cromossomo da linhagem de referência, seguida de uma “montagem individual” de cada cromossomo com o algoritmo IDBA-Hybrid (Montagem 01). Os *contigs* foram remontados pelo IDBA-Hybrid juntamente aos *reads* pareados sem a separação por cromossomo (Montagem 02). A montagem final utilizou o resultado da Montagem 02 como “*contigs* confiáveis” no programa SPAdes, em conjunto com os *reads* sem separação (Montagem final), com o objetivo de gerar *scaffolds* mais longos.

Para a anotação desse novo genoma, os transcritos de *S. cerevisiae* S288C do SGD (CHERRY et al., 2012) foram mapeados sobre tal genoma utilizando o GMAP (versão 2017-11-15, com os parâmetros `-n 1 --min-trimmed-coverage=0.70 --min-identity=0.90`) (WU; WATANABE, 2005) (evidência 1). Posteriormente, uma anotação *de novo* foi feita com o programa MAKER (CANTAREL et al., 2008) (evidência 2) e, por fim, uma curadoria manual, comparando as duas estratégias, foi realizada seguindo as regras: (I) Intersecções das evidências 1 e 2 foram consideradas corretas; (II) Discordâncias entre as coordenadas das evidências 1 e 2 e anotações não identificadas eram visualmente comparadas com o genoma de referência e suas coordenadas eram corrigidas ou removidas de acordo com o observado. Finalmente, anotações geradas pelo programa AGAPE (SONG et al., 2015) também foram utilizadas para correções nas coordenadas

dos genes e de suas respectivas *open reading frames* (ORFs) na anotação final. As sequências de proteínas foram traduzidas das regiões anotadas.

Adicionalmente, os elementos transponíveis (TEs) foram atualizados nas anotações de todas as linhagens. Para isso foi usado o programa RepeatMasker (parâmetros: -e crossmatch, -s, -lib repeat\_library.fasta, -pa 10, -a, e -gff) (TARAILO-GRAOVAC; CHEN, 2009). Primeiramente os elementos Ty foram obtidos no banco de dados RepBase (Update 22.09 – 20171013) (KOHANY et al., 2006). Em seguida, sequências do elemento Ty3 de *Saccharomyces paradoxus* (aqui denominadas TY3-1p\_I e TY3-1p\_LTR) também foram incorporadas ao banco de dados, pois também são encontradas no genoma de *S. cerevisiae* (CARR; BENSASSON; BERGMAN, 2012). Dessa maneira, foi criada uma biblioteca de elementos repetitivos para o RepeatMasker (repeat\_library.fasta). Finalmente, um *script* auxiliar em Perl do RepeatMasker (buildSummary.pl) foi utilizado para resumir as anotações.

As anotações de TEs encontradas em cada genoma foram curadas manualmente. Para este propósito, se dois TEs muito próximos (1 nucleotídeo de distância) tiverem o mesmo nome, foi decidido: 1 - assumi-los como apenas uma sequência expandindo uma única anotação se eles tiverem a mesma orientação; ou 2- reeditar a orientação da sequência menor se ambas tivessem orientações diferentes, expandindo uma única anotação. Essas modificações foram aplicadas sobre as anotações GFF. Além disso, “*Low-complexity*”, “*simple repeats*” e “*rRNA/Fungi*” não foram inseridos na atualização das anotações.

### 3.5. Análise de RNA-Seq e metabolomas: expressão diferencial , enriquecimento de GOs e anotações KEGG

Um total de 36 amostras de RNA (6 linhagens x 2 tratamento e controle x 3 replicatas) foram sequenciadas pela LcSciences (Texas, USA). As amostras passaram por depleção de ribossomos e foram sequenciadas utilizando o Illumina HiSeq 4000 (100 nts *paired-end* e tamanho de inserto de ~24~324pb).

Os *reads paired-end* foram processados com o Trimmomatic v0.36 (com os parâmetros ILLUMINACLIP illumina\_adapters.fa 2:25:10, HEADCROP 2, LEADING 3, TRAILING 3, SLIDINGWINDOW 4:30, MINLEN 95) (BOLGER; LOHSE; USADEL, 2014) e mapeados no genoma específico de cada linhagem utilizando o programa HISAT 2 v2.1.0 (com o parâmetro "--no-softclip") (KIM; LANGMEAD; SALZBERG, 2015). O número absoluto de *reads* sobre cada transcrito foi obtida com o programa Bedtools multiBamCov v2.26.0 (QUINLAN; HALL, 2010) e a expressão diferencial foi realizada com o DESeq2 (LOVE; HUBER; ANDERS, 2014). Foram considerados transcritos diferencialmente expressos aqueles com *false discovery rate* (FDR) <0,01, quando comparado tratamento vs controle. Adicionalmente, após o teste de normalidade de Shapiro-Wilk sobre a razão do número de genes sobre e sub expressos em  $\log_2$ , o teste T (não pareado) foi aplicado para determinar se havia diferenças entre os grupos HT e LT.

Todos os *fold-changes* genes diferencialmente expressos foram plotados em vias do KEGG (OGATA et al., 1999) utilizando o pacote Pathview (LUO; BROUWER, 2013). Para isso duas abordagens foram utilizadas: (I) uma abordagem na qual somente os genes diferencialmente expressos exclusivamente em HT e LT foram plotados, separadamente, e uma intersecção, na qual somente os genes diferencialmente expressos em ambos os grupos (HT e LT) foram plotados; e (II) uma abordagem geral

onde todos os genes, juntamente com os *fold-changes*, independente da significância estatística, foram plotados para todas as linhagens individualmente, a fim de obter uma visão geral sobre o estado da via.

Os genes diferencialmente expressos foram selecionados para análise de enriquecimento com o g:Profiler (REIMAND et al., 2016). Os termos de *gene ontology* (GO) enriquecidos foram sumarizados utilizando o REVIGO (SUPEK et al., 2011). Também, para acessar a significância estatística das vias do KEGG, todos os genes e *fold-changes* obtidos pelo DESeq2 foram submetidos a uma análise de enriquecimento via GAGE (LUO et al., 2009), foram considerados enriquecidas as vias com FDR <0,01.

Para os metabolomas o *software* ChromaTOF v. 4.51 (Leco Corp., St. Joseph, EUA) foi usado para executar a correção de *baseline* e também para exportar todos os arquivos MS no formato NetCDF. A detecção de pico, o tempo de retenção e correspondência de bibliotecas foram realizados usando o pacote TargetSearch (faixa de massa = 45.800, biblioteca da coluna MDN35 do banco de dados Gold, RI\_dev = 5, 2.5, 1.25, massas superiores = 20, massas excluídas = 147-149) (CUADROS-INOSTROZA et al., 2009). Metabólitos com contagem de massa  $\geq 3$  foram selecionados e identificados por meio da comparação de seus índices de retenção ( $\pm 2s$ ) e espectros (similaridade  $\geq 600$ ) com os compostos armazenados no banco de dados Golm-Metabolome (KOPKA et al., 2005). Se o *score* de massa de um metabólito era igual a 0 em  $\geq 3$  replicatas, foi pontuada como 0 em todas as outras replicatas. Depois, cada intensidade de metabólito foi normalizada de acordo com a equação 2 (eq. 2):

$$M_i = \frac{\left( \frac{M_{0i}}{\sum TIC_s} \right)}{M_{gi}} \quad (\text{eq. 2})$$

na qual  $M_i$  é a intensidade final do metabólito  $i$ ,  $M_{0_i}$  é a medida inicial do metabólito  $i$ ,  $TIC_s$  é a contagem total de íons da amostra que o  $M_i$  está presente;  $Mg_i$  é o peso em gramas da amostra  $i$ .

Na normalização das intensidades, os metabólitos foram avaliados quanto à abundância relativa dos mesmos, comparando tratamento vs controle para cada linhagem utilizando o MetaboAnalyst (CHONG; XIA, 2018). *Heatmaps (Pearson clustering)*, análise de componentes principais (PCA) e de mínimos quadrados parciais (análise discriminante PLS-DA) foram gerados para verificar o agrupamento dentro de amostras de controle e tratamento. O teste T (pareado) e o PLS-DA foram usados para calcular os metabólitos diferencialmente abundantes (DAMs) (valor  $Q \leq 0,05$ ), para cada linhagem, com base nas intensidades comparando tratamentos vs controle. Além disso, o  $\log_2$  *fold-change* (L2FC) foi usado para definir os metabólitos sobre e sub-abundantes.

A análise de conjuntos foi utilizada para avaliar os DAMs: 1 - os metabólitos presentes em todas as HTs e em todas as LTs foram comparados, mostrando o número de metabólitos presentes em ambos fenótipos quanto os fenótipos-exclusivos; 2 - os DAMs exclusivos de um fenótipo e presente em mais de uma linhagem foi separado em sobre ou sub-abundante; 3 - os DAMs presente nos fenótipos HT e LT também foi dividido em sobre e sub-abundante dentro de cada fenótipo (**Figura 11 suplementar; Apêndice 2**). Os identificadores dos metabólitos descritos no **apêndice 2** foram convertidos em um identificador do KEGG e seus L2FC foram plotados nos mapas juntamente com as informações dos transcritos seguindo o mesmo raciocínio dos dados de RNA-Seq.

### 3.6. Tratamento com espermidina

Para avaliar a influência da espermidina nas linhagens cultivadas sob a presença de etanol, todas as linhagens foram cultivadas em meio YPD à 30 °C e 130 RPM por 16

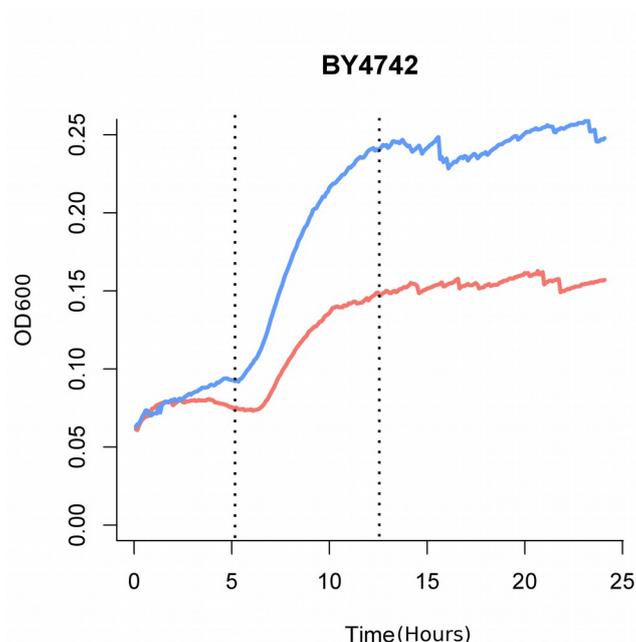
horas. Posteriormente, as linhagens foram cultivadas (em triplicata) em uma placa de 48 poços com meio YPD, etanol no volume máximo de tolerância para cada linhagem (**Tabela 2**) e espermidina (Sigma-Aldrich®, S2626) em uma concentração de 0,1 mM para o grupo tratamento, enquanto que no grupo controle foram adicionados apenas etanol e YPD. A OD<sub>600nm</sub> na placa de 48 poços foi lida em equipamento Epoch (BioTek®) a cada 5 minutos durante 24 horas com agitação contínua e temperatura de 30 °C.

Os resultados obtidos a partir da etapa supracitada foram ajustados em um modelo de curvas de crescimento populacional ao longo do tempo proposto por Pierre François Verhulst, cuja solução analítica é apresentada a seguir (**eq. 3**).

$$P(t) = M / (1 + ((1/P(0)) - (1/M)) * M * \exp(-k*t)) \quad (\text{eq. 3})$$

onde  $P$  representa a população ao longo do tempo  $t$ ,  $k$  é a taxa de crescimento e  $M$  representa o número máximo de indivíduos possíveis. Este modelo foi resolvido usando o MatLab e os resultados para  $k$  ficaram dentro do intervalo de confiança de 95%.

Como o objetivo era estudar a variação de  $k$  no tratamento vs controle, apenas a fase log e início da fase estacionária dos dados foram utilizadas para adequar-se ao modelo (**Figura 8**).



**Figura 8:** Representação do corte feito nos dados de crescimento para ajuste ao modelo. Aqui a linhagem BY4742 foi usada como exemplo. Isso foi feito para cada tensão e todos os resultados podem ser visualizados no tópico Resultados.

### 3.7. Rede global integrada e análises estatísticas

Para obter uma rede integrada específica para cada linhagem, foram obtidas redes de interação proteína-proteína (PPI) dos bancos de dados Biogrid (CHATRYAMONTRI et al., 2015) e Mint (LICATA et al., 2012). A rede regulatória (GRN) foi obtida do banco de dados *YTRP database* (YANG et al., 2014), a rede metabólica do banco de dados Reactome (FABREGAT et al., 2014), e uma rede de interação *long-non-coding* RNAs com proteínas (lnc-p) foi gerada pelo nosso grupo de pesquisa (dissertação de MARQUES LF relatório 10/11/2018 do Processo FAPESP 2017/14764-3). As redes obtidas foram filtradas para que somente interações de *S. cerevisiae* validadas experimentalmente fossem mantidas (com exceção da rede lnc-p). Adicionalmente, para redes PPI, as interações listadas no banco de dados Negatome (BLOHM et al., 2014) foram removidas.

A união das redes foi obtida com o pacote *igraph* (CSARDI; NEPUSZ, 2006) para formar uma única rede integrada (Uninet). A partir da Uninet, redes específicas para cada uma das linhagens analisadas foram construídas individualmente removendo os nós e as respectivas conexões de acordo com a presença/ausência de genes anotados em cada linhagem. Então, 13 características topológicas gerais foram obtidas, sendo elas *eccentricity*, *degree*, *eigenvectors*, *betweennes*, *closeness*, *page rank*, *strength*, *hub score*, *coreness*, *subgraph centrality*, *burstconstraint*, *transitivity* e *average nearest neighbor degree*, também através do pacote *igraph* (CSARDI; NEPUSZ, 2006). Após o teste de normalidade de Shapiro-Wilk sobre as características no ambiente de análise estatística R (R CORE TEAM, 2015), o teste de Kruskal-Wallis foi aplicado para determinar se as diferenças entre estas características eram significativas.

As 16 redes finais foram consideradas grafos não direcionados, uma vez que não é possível direcionar arestas em grandes redes, e pela ausência de informação de direção disponível para alguns dados aqui utilizados. Assim, em adição à avaliação estatística das 16 redes, utilizou-se a função DDC normalizada  $\square(k, k')$  (uma correlação normalizada grau-grau). Este método avalia a probabilidade de dois nós com um *degree*  $k$  e  $k'$  interagirem em comparação com o que seria encontrado em uma rede aleatória com a mesma distribuição de  $k$  não correlacionados. Desse modo é possível efetuar uma caracterização “microscópica” dessas redes, e avaliar até que ponto elas diferem de um modelo aleatório (FERNANDES et al., 2010).

### 3.8. Machine Learning aplicado a redes biológicas

As conexões da rede Uninet foram modificadas com base nos resultados da expressão diferencial com o objetivo de criar duas redes distintas para cada linhagem,

representando as conexões ativas no estado de controle e no estado de tratamento. O procedimento foi realizado da seguinte forma:

1 – Para cada linhagem os genes foram separados em ativos (genes diferencialmente expressos com *fold-change* positivo e FDR <0.01), inativos (genes diferencialmente expressos com *fold-changes* negativos e FDR <0.01) e neutros (genes sem diferença de expressão significativa);

2 – Genes sem expressão (0 *read counts*) foram removidos da rede. Assim somente as conexões dos genes expressos foram mantidas na Uninet;

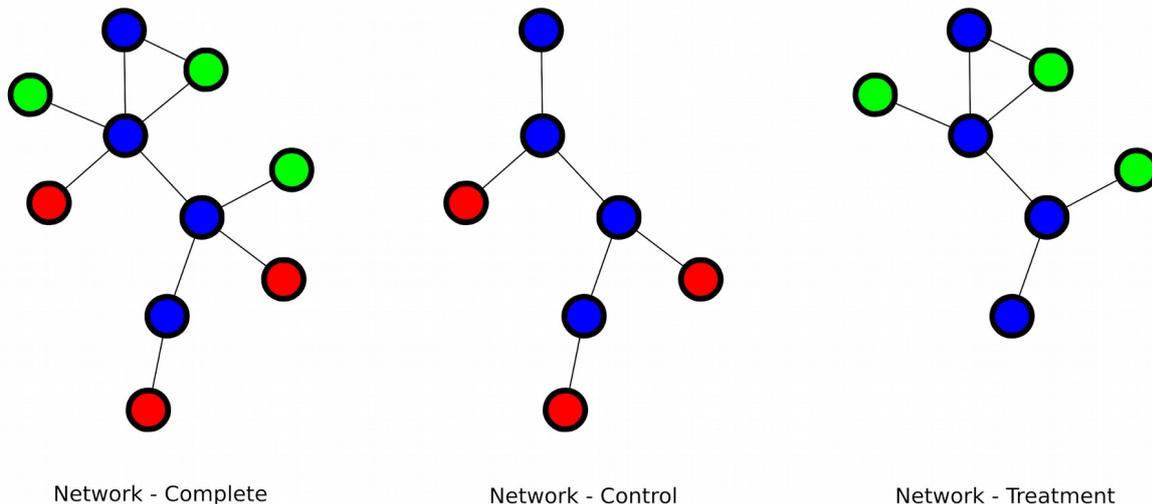
3 – **“Rede controle”**: genes ativos tiveram suas conexões (arestas) removidas da Uninet (**Figura 9 – Network - Control**);

4 – **“Rede tratamento”**: os genes inativos no tratamento (concentração de etanol máxima suportada pela linhagem) são removidos da Uninet, e os genes ativos e suas respectivas conexões (arestas) são mantidos (**Figura 9 – Network treatment**);

5 – As matrizes de adjacência de cada rede gerada (“rede controle” e “rede tratamento”) foram obtidas para cada linhagem. Nestas matrizes cada elemento  $a_{ij}$  corresponde à interação entre os genes  $i$  e  $j$ , podendo esse elemento ser 0 para indicar que o gene  $i$  não possui conexão com o gene  $j$ , ou 1 para indicar que há uma conexão ativa entre os dois genes;

6 – Para cada matriz uma coluna adicional com um rótulo (“controle” ou “tratamento”, de acordo com a respectiva matriz) foi adicionado para cada linha gerando a “Matriz CT”. Adicionalmente uma matriz alternativa com os rótulos “HT” ou “LT” foi criada, gerando a “Matriz HTLT”;

7 – As matrizes “CT” de todas as linhagens foram unidas. Cada linha representando uma instância (um gene) e cada coluna os atributos (genes com os quais o gene da instância tem conexões ativas). Como o número de instâncias era o mesmo em todas as matrizes, os dados podem ser considerados balanceados. Desse modo é esperado que, se uma dada conexão é muito frequente em uma classe específica, a mesma será utilizada pelo *Machine Learning* (ML) para classificação. O mesmo procedimento foi aplicado para as matrizes “HTLT”.



**Figura 9:** Geração das redes de controle e tratamento. **Vermelho:** genes inativos (genes diferencialmente sub expressos); **Verde:** genes ativos (diferencialmente sobre expressos); **Azul:** genes neutros (genes sem expressão diferencial). **Network – Complete:** rede integrada completa; **Network – Control:** rede com as conexões ativas no estado de controle; **Network – Treatment:** rede com as conexões ativas no estado de tratamento.

A matriz resultante foi utilizada para o ML com o algoritmo de árvore de decisão J48 implementado no *software* Weka (HALL et al., 2009). Assim, ao invés de gerar um modelo preditivo, o objetivo é obter uma árvore de decisão pequena na qual seja possível observar quais genes são mais relevantes na rede (em termos de estarem conectados ou não) para classificar as linhagens em estado de controle e tratamento ou em termos de altamente ou pouco tolerantes ao etanol.

Adicionalmente às matrizes de adjacência binária (representando as conexões entre os genes por meio de “0” ou “1”) também foi testada uma matriz com as relações de expressão discretizada entre os pares de genes. Nestas matrizes, cada elemento  $a_{ij}$  corresponde à interação entre os genes  $i$  e  $j$ , podendo esse elemento ter um valor de “A”, “B”, “C”, “D”, “E” ou “F”, de acordo com a relação da expressão diferencial dos genes  $i$  e  $j$ . A discretização dos valores de cada um dos elementos permitiu atribuir letras de acordo com o perfil de expressão diferencial da seguinte maneira: “A” quando os dois genes conectados não eram diferencialmente expressos; “B” quando um gene de uma conexão foi sobre-expresso e o outro não apresentava expressão diferencial; “C” quando um gene de uma conexão foi sub-expresso e o outro não apresentava expressão diferencial; “D” quando os dois genes conectados eram diferencialmente sobre-expressos; “E” quando os dois genes conectados eram diferencialmente sub-expressos; e “F” quando os dois genes conectados diferiram em sua expressão diferencial (um é sobre e o outro é sub-expresso).

## 4. Resultados

### 4.1. O nível de tolerância ao etanol

As linhagens selecionadas variam de 18-30% (v/v) de tolerância máxima de EtOH (**Tabela 2**). As linhagens determinadas como HT são: BMA64-1A, BY4742 e X2180-1A; e as determinadas como LT são: S288C, BY4741 e SEY6210 (ALMEIDA, 2017). Além disso o *mating types* (MAT), que é o *locos* responsável para determinar a compatibilidade entre duas células haplóides de *S. cerevisiae* para que estas formem uma célula diploide (*mating*) (BUTLER, 2010), parece não influenciar a tolerância em virtude de ambas variantes, “a” e “alpha”, estarem presentes tanto para HTs como para LTs.

**Tabela 2:** Agrupamento das linhagens estudadas e suas tolerâncias máximas ao EtOH. <sup>1</sup>: linhagem com o genoma sequenciado nesse trabalho; **EtOH Tol.:** Tolerância máxima ao etanol. **Fonte:** (ALMEIDA, 2017).

<b>Linhagem</b>	<b>EtOH Tol.</b>	<b>Cluster</b>	<b>Ploidia (MAT <i>locos</i>)</b>
BMA64-1A <sup>1</sup>	30	HT	a
BY4742	26	HT	alpha
X2180-1A	24	HT	a
S288c	20	LT	alpha
SEY6210	20	LT	alpha
BY4741	22	LT	a

### 4.2. Montagem e anotação genômica

Um total de 6.464.944 *reads* foram obtidos do sequenciamento do genoma da linhagem BMA64-1A. Após o filtro de qualidade, 6.410.152 *reads* (com aproximadamente 250 nts de comprimento) foram obtidos. As estratégias de montagem diferentes para a linhagem BMA64-1A resultaram em 196 montagens e as métricas gerais podem ser vistas na **Tabela 3**. Uma vez que foi utilizada uma equação para compilar múltiplas métricas, algumas vezes a montagem escolhida não possuía os melhores valores individualmente.

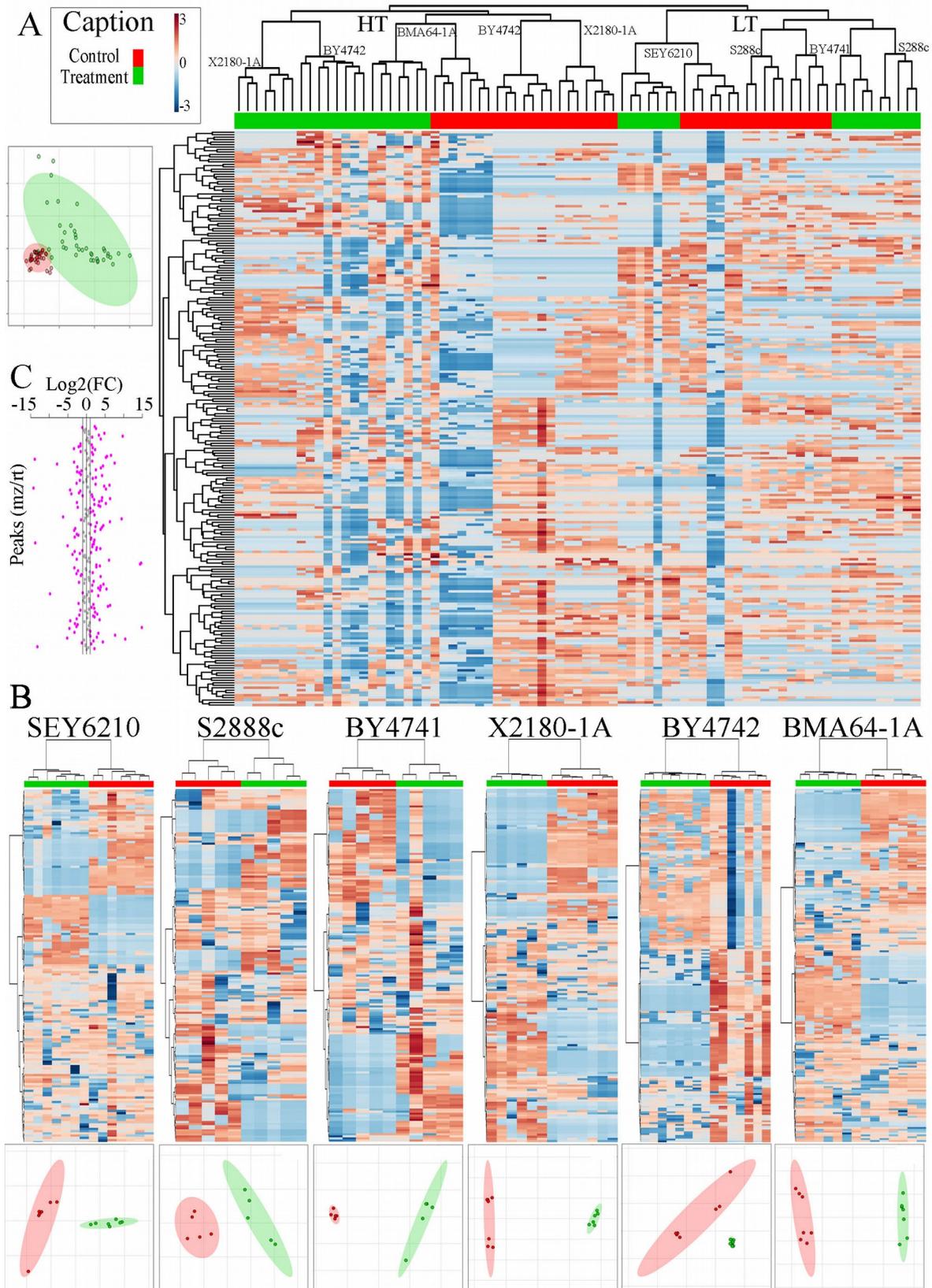
**Tabela 3:** Resumo das métricas de montagem da linhagem BMA64-1A considerando 196 montagens. **Min:** valor mínimo; **Máx:** Valor máximo; **Selec:** métricas da montagem selecionada; **F/M:** Faixas/Métricas; **CS:** número de *contigs/scaffolds*; **TT:** Tamanho total; **GR:** Genes provavelmente recuperados; **SM:** *Assembly Score* da montagem.

F/M	N50	CS	GC%	TT	GR	SM
Min	1.437	17	37.93	126.580	1.744	0.054
Máx	924.585	27.639	38.34	12.162.499	5.872	0.63
Selec	202.259	683	38.06	11.880.801	5.668	0.29

A montagem selecionada gerou um genoma com tamanho total de 11.880.801 nts, alocado em 683 *scaffolds*, com uma cobertura de 134,88x. O processo de anotação resultou em 6.551 genes codificantes dos quais somente 36 (aqui atribuídos como supostos novos genes) não possuem similaridade com o genoma de referência. Logo, a grande maioria dos genes codificantes são congruentes com os do genoma da linhagem de referência S288C. Os supostos 36 prováveis genes novos aqui reportados poderão ser adicionados ao pan-genoma (o conjunto de todos os genes conhecidos para a espécie) de *S. cerevisiae*, o qual continua crescendo ao longo dos anos. Contudo, essa estratégia permitiu construir um genoma comparável com *drafts* disponíveis nos bancos de dados e, de fato, o método montou um número de *scaffolds* menor do que todos os genomas já montados para *S. cerevisiae*, sendo a linhagem RM11-1A (disponível no SGD) (ver (SONG et al., 2015)) e S288C as exceções. Além disso, obteve-se grandes regiões genômicas com grupos de genes, demonstrado pelo maior tamanho total e N50 quando comparadas com as montagens geradas pelo *pipeline* AGAPE.

#### 4.3. Análise de abundância diferencial de metabólitos

Comparando os tratamentos vs controle, é possível observar divergências na abundância de metabólitos comparando ambas as condições (**Figura 10**). Assim, as comparações relativas aos metabólitos diferencialmente abundantes (DAM) demonstram ser viáveis para as nossas amostras (**Figura 10A-B**). O número de DAMs por linhagem varia de 101 (SEY6210) a 182 (BY4742) metabólitos, sem qualquer tendência para DAMs sobre ou sub abundantes, comparando todas as amostras disponíveis no **Apêndice 2**.



**Figura 10:** Visão geral da análise metabólica. **A:** *heatmap* e PLS-DA comparando todas as amostras; **B:** *heatmap* e PCA comparando amostras dentro de cada linhagem; **C:** histograma de DAMs considerando todas as amostras. Todos os *heatmaps* foram realizados usando correlação de Pearson. Tanto no PCA como na PLS-DA, os eixos X e Y são os componentes principais 1 e 2, respectivamente. As cores dos controles e tratamento de todas as parcelas estão presentes na caixa “caption”.



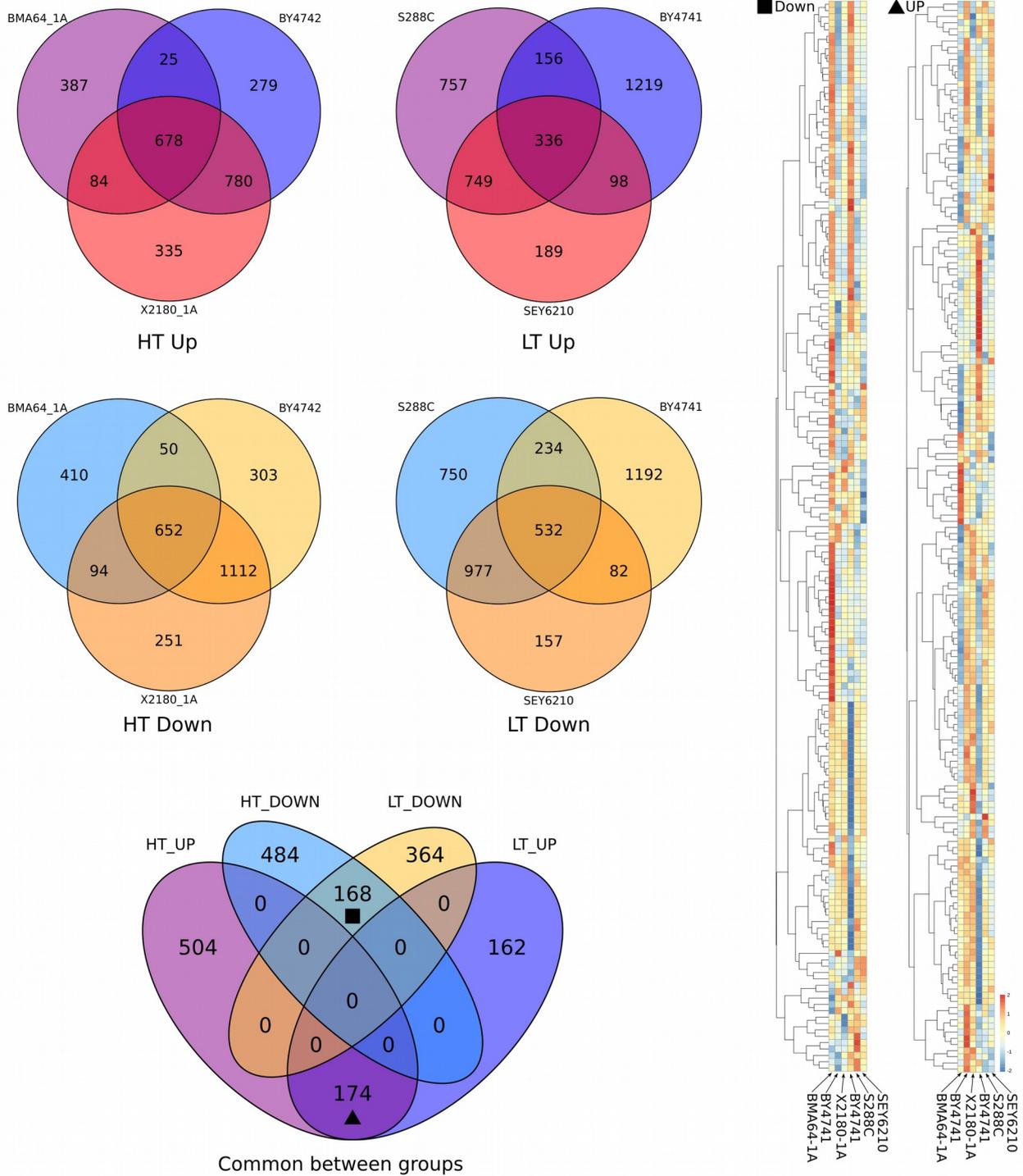
#### 4.4. Análise de expressão diferencial

Os experimentos de expressão diferencial mostraram que em média 50,31% dos genes diferencialmente expressos estão sobre expressos em HT e 47,76% em LT, enquanto que 49,68% estão sub expressos em HT e 52,23% em LT (**Tabela 4**). Ademais, após determinar que a razão da variação dos genes sobre e sub expressos era normal (HTs:  $W = 0,9939$ ,  $p\text{-value} = 0,8507$  e LTs:  $W = 0,99914$ ,  $p\text{-value} = 0,9441$ ) foi observado que estas razões são significativamente diferentes considerando as linhagens HTs e LTs ( $t = 3,3179$ ,  $p\text{-value} = 0,0442$ ). Isso indica que as linhagens HTs tendem a conter mais genes sobre expressos do que as linhagens LTs.

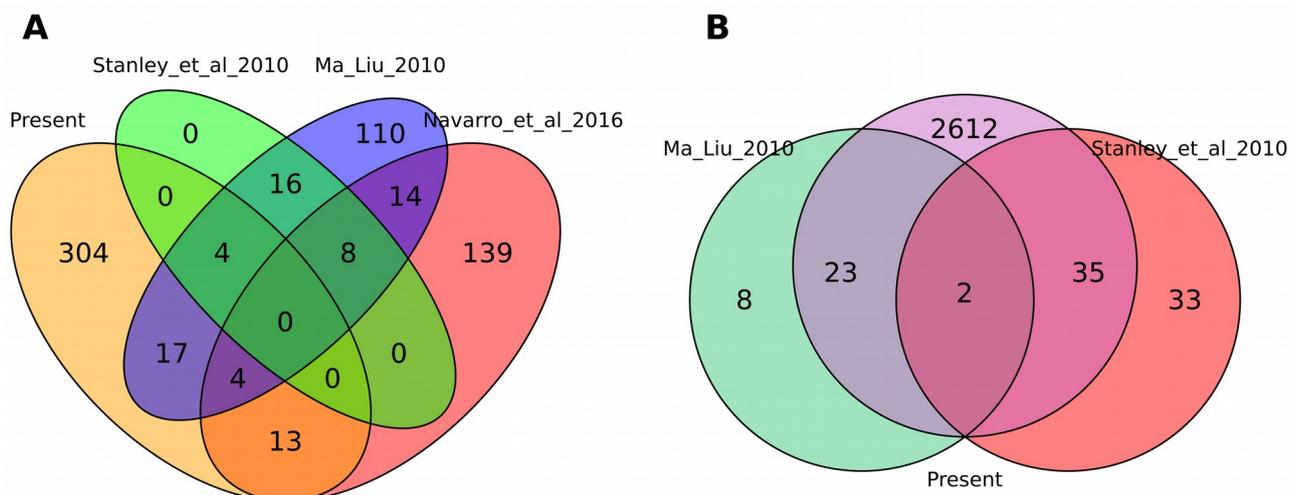
A comparação qualitativa dos genes diferencialmente expressos entre as linhagens HT e entre as LT reportam um total de 1.330 e 868 genes diferencialmente expressos, respectivamente (soma dos centros dos diagramas de Venn da **Figura 12**). As linhagens HT mostraram um total de 678/652 genes sobre/sub expressos, enquanto que as LT mostraram um total de 336/532 genes sobre/sub expressos (**Figura 12**). Ainda, 174/168 genes são sobre/sub expressos em comum entre os grupos HT e LT (**Figura 12**). Ao comparar os genes diferencialmente expressos da intersecção entre linhagens HT e LT com estudos anteriores, é possível observar que 38 genes já haviam sido citados anteriormente e 304 foram exclusivos deste estudo (**Figura 13A**).

**Tabela 4:** Número de genes diferencialmente expressos por linhagem. **Group:** grupo, HT para altamente tolerante ao etanol e LT para pouco tolerante ao etanol; **Up:** número de genes sobre expressos no tratamento em comparação ao controle; **Down:** número de genes sub expressos no tratamento em comparação ao controle; **Total DE Genes:** Número total de genes diferencialmente expressos; **% UP:** porcentagem de genes sobre expressos; **% DOWN:** porcentagem de genes sub expressos.

Strain	Group	Up	Down	Total DE Genes	% UP	% DOWN
BMA64-1A	HT	1319	1240	2559	51.54	48.45
BY4742	HT	2050	2117	4167	49.19	50.80
X2180-1A	HT	2144	2126	4270	50.21	49.78
S288C	LT	2277	2552	4829	47.15	52.84
BY4741	LT	2116	2256	4372	48.39	51.60
SEY6210	LT	1640	1795	3435	47.74	52.25



**Figura 12:** Comparação quantitativa e qualitativa dos genes diferencialmente expressos entre as linhagens analisadas. **HT Up:** genes sobre expressos em linhagens HT; **HT Down:** genes sub expressos em linhagens HT; **LT Up:** genes sobre expressos em linhagens LT; **LT Down:** genes sub expressos em linhagens LT; **Common between groups:** genes sobre e sub expressos a partir do núcleo dos diagramas superiores; **Quadrado:** heatmap dos genes sub-expressos em todas as linhagens; **Triângulo:** heatmap dos genes sobre-expressos em todas as linhagens.



**Figura 13:** Diagramas comparando genes e termos GO encontrados em diferentes estudos. **A:** **Stanley et al 2010:** genes citados em (STANLEY et al., 2010a); **Ma liu 2010:** genes citados em (MA; LIU, 2010); **Navarro et al 2016:** genes citados em (NAVARRO-TAPIA et al., 2016); **Present:** genes citados neste estudo. **B:** **Liu and MA 2010:** Termos GOs de (MA; LIU, 2010); **Stanley et al 2010:** Termos GOs de (STANLEY et al., 2010a); **Present:** Todos os termos Gos únicos encontrados neste estudo para todas as linhagens com valor de  $p < 0.05$  segundo o g:Profiler.

Alguns dos principais genes relacionados aos estressores básicos durante o processo de fermentação recentemente publicado (obtidos de (AUESUKAREE, 2017)) estavam dentre os genes diferencialmente expressos aqui encontrados (**Tabela 5**). É interessante notar que:

1. há um maior número de genes diferencialmente sobre expressos em linhagens HT;
2. alguns genes relacionados a resposta ao estresse oxidativo e calor são comumente expressos em ambos os grupos de linhagens;
3. os genes que respondem ao estresse osmótico que ativam as vias *mitogen-activated protein kinase (MAPK)* na levedura estão em sua maioria sub expressos (PBS2, MSB2, SSK22, GPD2, STE20);
4. *heat shock proteins (HPSs)* que auxiliam na estruturação, recuperação e degradação de proteínas estavam sobre expressas em linhagens HT;

5. genes da família SOD, que convertem ânions superóxidos em peróxido de hidrogênio para que sejam então reduzidos a água e oxigênio por catalases (como a CTA1) ou glutathione peroxidases (como a GPX1), também estavam sobre expressos em linhagens HT;
6. enzimas que controlam o estado de redução de proteínas como a TRX2 (citoplasmática) e TRX3 (Mitocondrial) também foram observadas;
7. a GSH1 está envolvida na biossíntese de glutathione, que pode agir como antioxidante, estava sobre expressa em linhagens HT;
8. os genes TPS1 e TPS2, que produzem trealose (um dissacarídeo importante para manter a estabilidade de proteínas), são sensíveis ao etanol e também estavam super expressos em linhagens HT;
9. os genes SS3 e SS4, membros da família de HSP70 e que atuam no redobramento de proteínas no estresse por etanol e calor, estavam sobre expressos em todas as linhagens, com exceção de BY4741;
10. os genes HTX1 e HTX2, responsáveis pelo transporte de glicose, estavam sub expressos em todas as linhagens, com exceção de HTX1, o qual só não estava reprimido nas linhagens X2180-1A e BMA64-1A;
11. dentre as álcool desidrogenases, que atuam na conversão de etanol para acetaldeído e também na reação reversa, a ADH1 estava sobre expressa somente em BMA64-1A. A ADH2 estava sobre expressa em BMA64-1A e BY4741 e não foi diferencialmente expressa em X2180-1A. No entanto, a ADH5 (paróloga de ADH1) e ADH3 estavam sobre expressas em todas as linhagens. Além disso, algumas enzimas importantes na conversão de compostos como a aldeído desidrogenase citoplasmática (ALD6) estava inativa em todas as linhagens, enquanto que a mitocondrial (ALD4) estava ativa em todas as linhagens;

12.a enzima ACS1 (enzima converte o acetaldeído em acetil-CoA) estava sobre expressa em todas as linhagens com exceção da BY4741.

A expressão diferencial dos TEs indica que a maioria deles são retrotransposons LTR. As linhagens HT também apresentam a maioria dos TEs sub expressos em comparação com as LTs. Ademais, LTs têm mais TEs expressos diferencialmente do que HTs.

A análise de termos GO relacionados aos genes diferencialmente expressos reportou 2.672 termos. Ao comparar os termos encontrados com estudos anteriores, foi possível observar que 23 termos já haviam sido representados por Ma e Liu em 2010, e 35 haviam sido encontrados nos estudos de Stanley e colaboradores também em 2010 (**Figura 13B**). Somente os termos GO:0005739 (*mitochondrion*) e GO:0007005 (*mitochondrion organization*) foram comuns entre os três estudos.

**Tabela 5:** Genes diferencialmente expressos comumente relacionados a alguns estressores. **Stressor:** tipo de estressor, **HT:** refere-se à linhagens HT, **LT:** refere-se à linhagens LT; **HTLT:** refere-se à genes comuns entre as linhagens HT e LT; **UP:** refere-se à genes sobre expressos; **DOWN:** refere-se à genes sub expressos; \*: não foram encontrados genes diferencialmente expressos nesta categoria.

Stressor	HT UP	HT DOWN	LT UP	LT DOWN	HTLT UP	HTLT DOWN
osmotic	GPD1 (YDL022W)	PBS2 (YJL128C), MSB2 (YGR014W), GPP2 (YER062C)	*	SSK22 (YCR073C), GPD2 (YOL059W)	*	STE20 (YHL007C)
heat	SSE2 (YBR169C), KAR2 (YJL034W), ECM10 (YEL030W), CDC37 (YDR168W), AHA1 (YDR214W), CPR6 (YLR216C)	*	*	SSE1 (YPL106C)	SSC1 (YJR045C), STI1 (YOR027W)	*
oxidative	SOD1/2 (YJR104C/sod1p, YHR008C/sod2p), CTA1 (YDR256C), GPX1 (YKL026C), AHP1 (YLR109W), PRX1 (YBL064C), TRX2/3 (YGR209C, YCR083W), GSH1 (YJL101C)	*	*	*	NTH1 (YDR001C), CTT1 (YGR088W), TSA2 (YDR453C), TRR2 (YHR106W)	*
etanol	HSP12/26/30/82/104 (YFL014W, YBR072W, YCR021C, YPL240C, YLL026W), SSA3/4 (YBL075C, YER103W), TPS1/2 (YBR126C, YDR074W)	MSN2 (YMR037C), TRP2/5 (YER090W, YGL026C), TAT2 (YOL020W)	*	*	*	HSP78 (YDR258C), TRP3 (YKL211C)

Após a sumarização dos termos GO, sete termos foram comuns entre os genes sobre-expressos nas três HTs sendo eles *eisosome assembly*, *coenzyme metabolism*, *catabolism*, *hydrogen peroxide metabolism*, *generation of precursor metabolites and energy*, *cofactor metabolism*, *carbohydrate metabolism*; quanto aos termos referentes aos genes sub-expressos nas três HTs, somente o termo *cell wall macromolecule metabolism* foi reportado. Quanto aos termos sobre-expressos nas três LTs, somente o termo *response to stimulus* foi reportado, porém, quando consideradas somente as linhagens S288C e SEY6210, 11 termos são encontrados (*import into cell*, *catabolism*, *pyridine-containing compound metabolism*, *response to stimulus*, *autophagy*, *hydrogen peroxide metabolism*, *cell communication*, *generation of precursor metabolites and energy*, *carbohydrate metabolism*, *cofactor metabolism*, *coenzyme metabolism*); situação similar ocorre com os termos relacionados aos genes sub-expressos, no qual não há termos em comum entre as três LTs ao passo que quatro são comuns entre S288C e SEY6210 (*localization within membrane*, *glycosylation*, *protein folding*, *methylation*).

Entre os diferentes grupos de linhagens, seis termos relacionados aos genes sobre-expressos em HTs também estão relacionados à genes de LTs (aqui considerando somente as linhagens S288C e SEY6210) sendo eles *coenzyme metabolism*, *catabolism*, *hydrogen peroxide metabolism*, *generation of precursor metabolites and energy*, *cofactor metabolism*, *carbohydrate metabolism*. O termo *eisosome assembly* é o único termo exclusivo de HT e os termos *import into cell*, *pyridine-containing compound metabolism*, *response to stimulus*, *autophagy* e *cell communication* são termos exclusivos para LTs.

Quando considerados os genes diferencialmente expressos em comum entre os grupos HT e LT, podemos observar uma representação diferente quando obtemos as intersecções por linhagem (**Tabela 6**).

**Tabela 6:** Processos biológicos obtidos pelo REVIGO em grupos HT e LT. **Sobre representados:** termos representados para os genes sobre expressos; **Sub representados:** termos representados para os genes sub expressos; **Intersecção HTLT:** termos para os genes na intersecção dos grupos HT e LT; **Exclusivos HT:** termos exclusivos para os genes do grupo HT; **Exclusivos LT:** termos exclusivos para os genes do grupo LT.

Sobre representados	Sub representados
<b>Intersecção HTLT</b>	
energy derivation by oxidation of organic compounds cofactor metabolism carbohydrate metabolism generation of precursor metabolites and energy cellular carbohydrate metabolism mitochondrion organization response to oxygen-containing compound	U4 snRNA 3prime-end processing ribonucleoprotein complex biogenesis
<b>Exclusivos HT</b>	
energy derivation by oxidation of organic compounds response to toxic substance cofactor metabolism cellular carbohydrate metabolism carbohydrate metabolism generation of precursor metabolites and energy oxidoreduction coenzyme metabolism pyridine-containing compound metabolism protein refolding	double-strand break repair DNA conformation change chromosome segregation biological regulation cellular component biological regulation organization or biogenesis cellular process
<b>Exclusivos LT</b>	
transmembrane transport cell growth	tRNA metabolism RNA transport ribonucleoprotein complex biogenesis cellular process

A análise de enriquecimento encontrou vias enriquecidas para todas as linhagens, com exceção da linhagem BY4741, a qual não possuía vias com FDR <0,01. Dentre as vias analisadas, aquelas relacionadas aos mecanismos essenciais na célula possuem um grande número de genes diferencialmente expressos. Estas vias incluem o ciclo celular, mecanismos de tradução e transporte de RNA e metabolismo de compostos e energia (**Tabela 7**). Dentre as particularidades observadas: 1 - o ciclo do citrato estava com seus genes sobre representados em todas as linhagens; 2 - o de fosforilação oxidativa só não estava representado na linhagem SEY6210; 3 - a via do proteassomo estava sobre

representada somente na linhagem BMA64-1A; 4 - a via do ciclo celular estava sub representada em todas as linhagens HT; 5 - as demais vias estavam sub representadas nas demais linhagens.

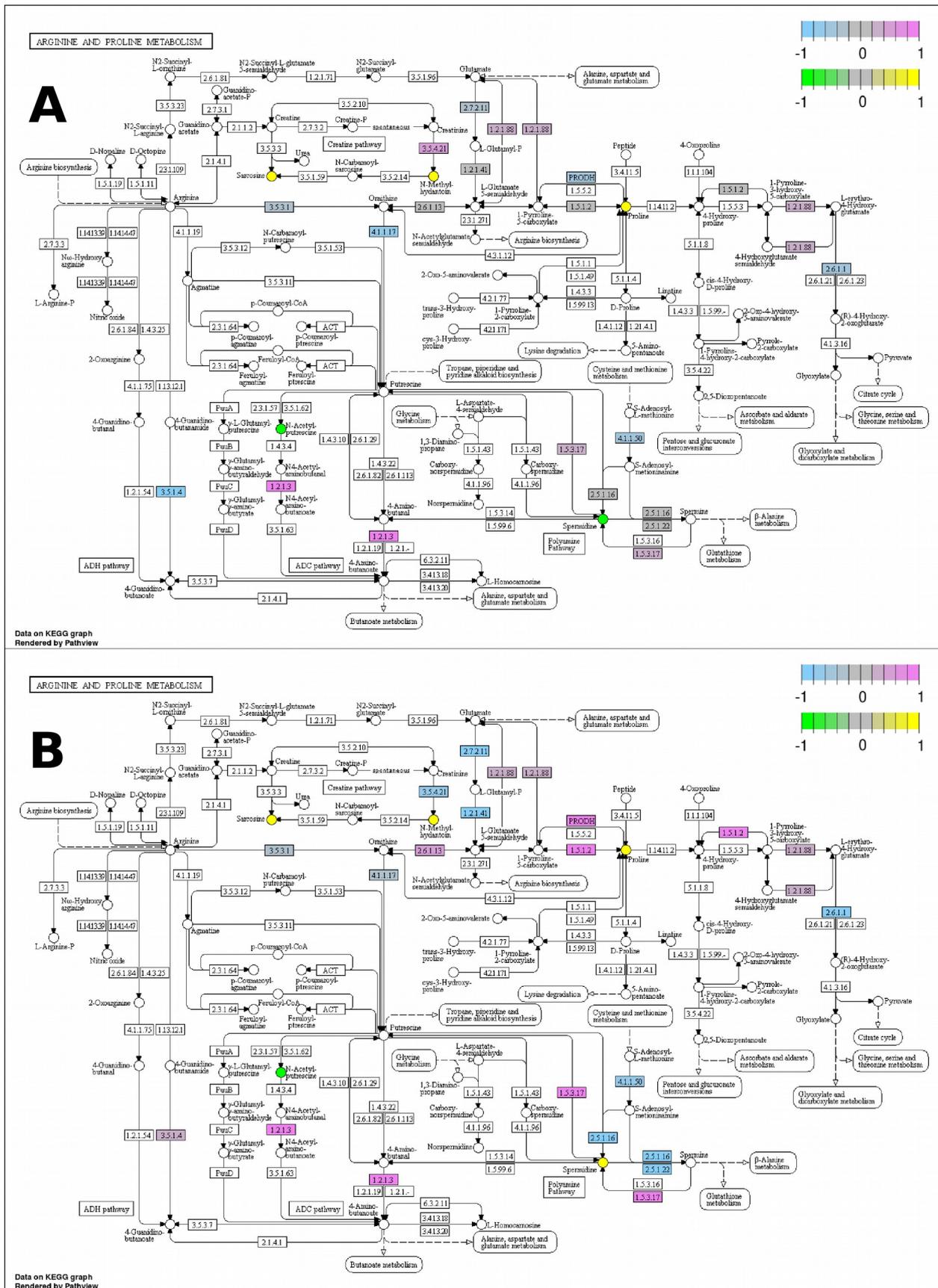
**Tabela 7:** KEGG *pathways* significativas detectadas pelo método GAGE. **Strain:** Linhagem; **FDR:** valor de p ajustado pela *False Discovery Rate*; **KeggName:** Número de acesso da via; **Pathway:** nome da via; \*: Mapas onde a spermidina está presente, baseado no resultado do metaboloma.

Strain	p.val	KeggName	Pathway
BMA64-1A	6.12097194656886e-05	sce03050	Proteasome
BMA64-1A	0.000121834426662788	sce00020	Citrate cycle (TCA cycle)
BMA64-1A	0.00013585969635272	sce00190	Oxidative phosphorylation
BMA64-1A	7.31643819776325e-09	sce04111	Cell cycle - yeast
BY4742	8.70419989854964e-08	sce00020	Citrate cycle (TCA cycle)
BY4742	2.01498358539526e-05	sce00190	Oxidative phosphorylation
BY4742	3.48740519371016e-32	sce03010	Ribosome
BY4742	6.70567447513877e-14	sce03008	Ribosome biogenesis in eukaryotes
BY4742	4.19347315634964e-08	sce03013	RNA transport
BY4742	3.44976440792473e-07	sce00240	Pyrimidine metabolism
BY4742	8.96963962561768e-07	sce00230	Purine metabolism
BY4742	1.44946367974482e-06	sce03020	RNA polymerase
BY4742	1.45610743421563e-05	sce04111	Cell cycle - yeast
BY4742	0.000277050571493239	sce00100	Steroid biosynthesis
BY4742	0.000813264978481155	sce00970	Aminoacyl-tRNA biosynthesis
BY4742	0.00120318185365423	sce03030	DNA replication
X2180-1A	5.68980825512778e-08	sce00020	Citrate cycle (TCA cycle)
X2180-1A	1.4833971592175e-05	sce00190	Oxidative phosphorylation
X2180-1A	1.04258525356658e-30	sce03010	Ribosome
X2180-1A	6.50443381507132e-14	sce03008	Ribosome biogenesis in eukaryotes
X2180-1A	2.31942125165522e-08	sce03013	RNA transport
X2180-1A	2.83068709892314e-06	sce03020	RNA polymerase
X2180-1A	1.57552617280349e-05	sce00240	Pyrimidine metabolism
X2180-1A	4.87375232012489e-05	sce00230	Purine metabolism
X2180-1A	0.000109295922857851	sce04111	Cell cycle - yeast
S288C	5.1921057074566e-07	sce00190	Oxidative phosphorylation
S288C	1.74715629951367e-06	sce00020	Citrate cycle (TCA cycle)
S288C	9.3739695305936e-69	sce03010	Ribosome
S288C	1.71149857058846e-12	sce03008	Ribosome biogenesis in eukaryotes
S288C	2.13463654055575e-07	sce03013	RNA transport
S288C	4.22224093152596e-07	sce03020	RNA polymerase
S288C	1.80197194053863e-06	sce00230	Purine metabolism
S288C	2.2315303226321e-06	sce00240	Pyrimidine metabolism
S288C	0.00061288354461916	sce00100	Steroid biosynthesis
SEY6210	9.11500500273983e-06	sce00020	Citrate cycle (TCA cycle)
SEY6210	5.58078784668617e-43	sce03010	Ribosome
SEY6210	5.63263026090742e-12	sce03008	Ribosome biogenesis in eukaryotes
SEY6210	1.024814414965e-06	sce00240	Pyrimidine metabolism
SEY6210	1.04461756849023e-06	sce00230	Purine metabolism
SEY6210	1.23795537618866e-06	sce03013	RNA transport

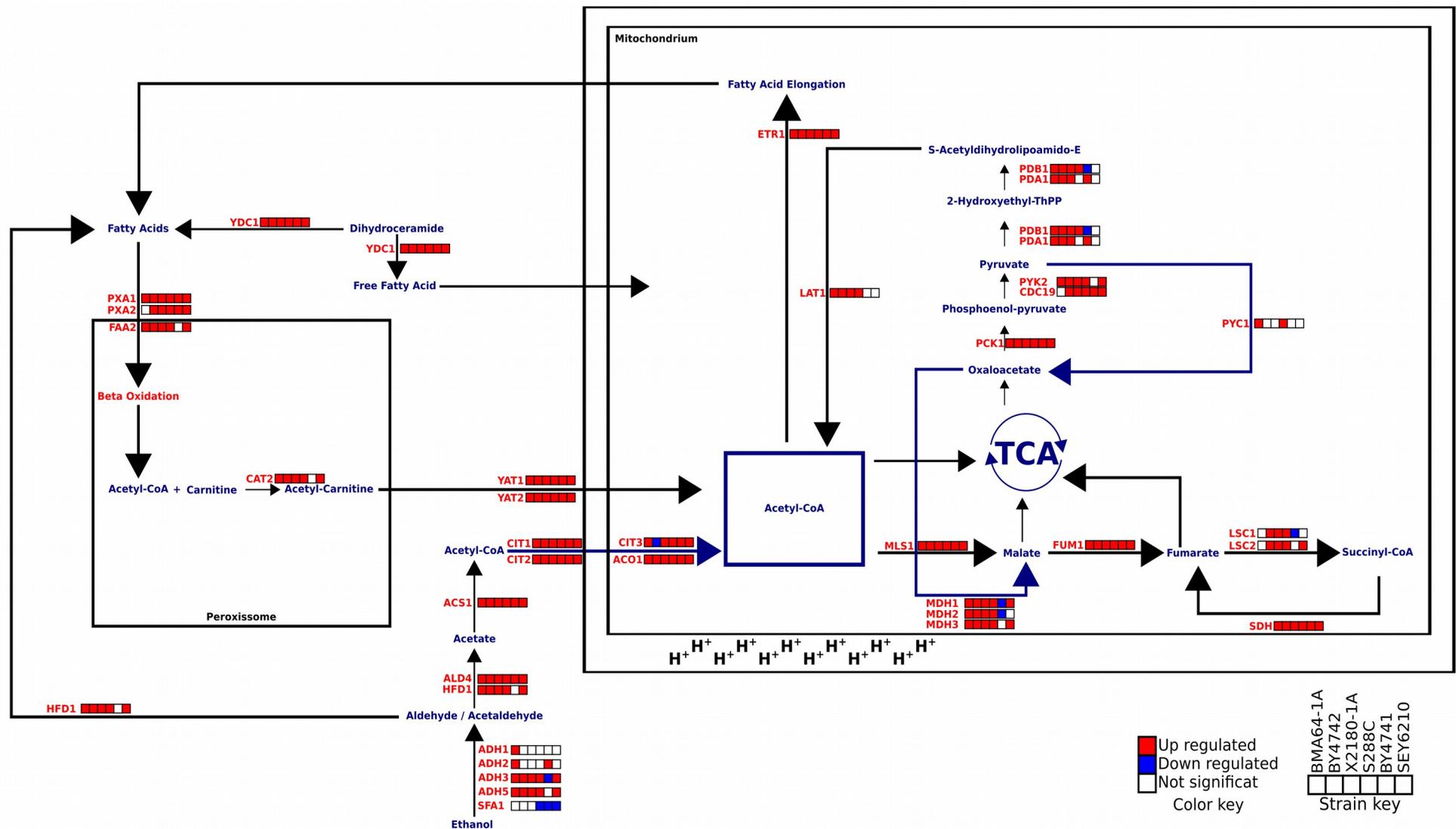
SEY6210	4.7975532816262e-06	sce03020	RNA polymerase
SEY6210	0.00079563630876036	sce00513	Various types of N-glycan biosynthesis
*	*	sce00330	Arginine and proline metabolism
*	*	sce00480	Glutathione metabolism
*	*	sce00410	beta-Alanine metabolism
*	*	sce02010	ABC transporters

A espermidina parece exercer um papel importante no crescimento celular e a prolina na tolerância ao etanol (**Tabela 7**), assim, a via dos metabolismos de arginina e prolina foi plotada para verificação do estado geral da espermidina e dos genes na via metabólica (**Figura 14**).

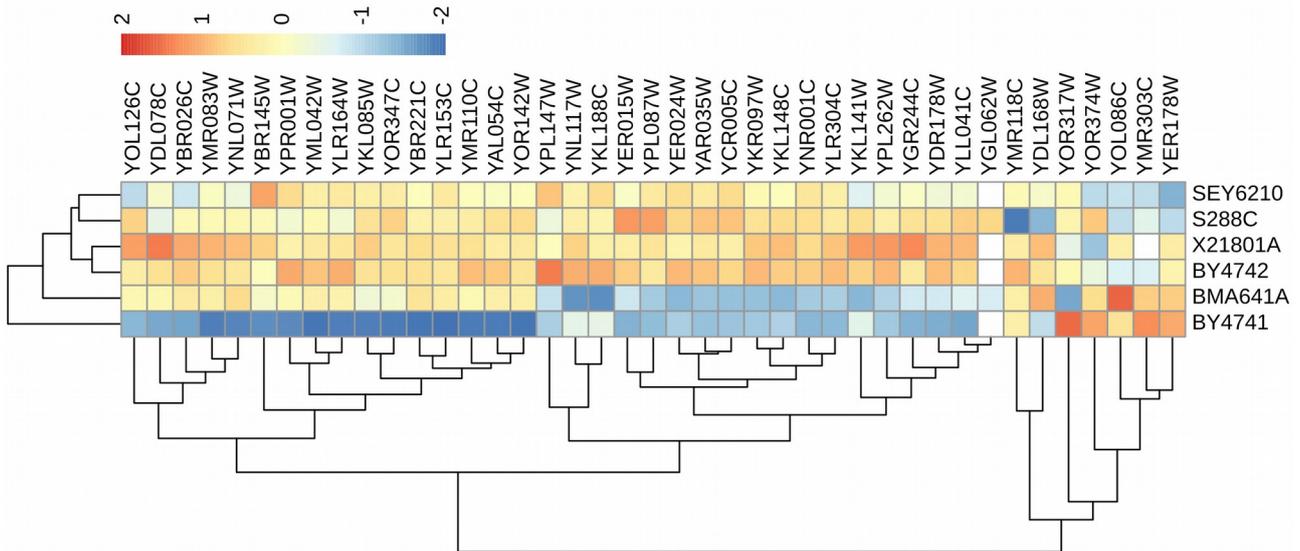
Em uma visão geral, a análise das informações sobre os genes diferencialmente expressos permitiu inferir um modelo de tamponamento de etanol baseado na via de *diauxic shift* (PISKUR et al., 2006) na levedura (**Figura 15**). Ademais, ao observar o agrupamento destes genes pelo nível de expressão, pode-se observar um cluster HT e um LT, sendo que apenas as linhagens BMA64-1A e BY4741 estão fora dos agrupamentos esperados (**Figura 16**).



**Figura 14:** Via do metabolismo de arginina e prolina. **A:** HT, BMA64-1A; **B:** LT, S288C; **Gradiente azul-rosa:** Transcritos diferencialmente expressos; **Gradiente verde-amarelo:** Metabolitos diferencialmente abundantes. A espermidina está reportada na porção inferior e a prolina no canto superior direito em ambas as figuras.



**Figura 15:** Modelo proposto para o fluxo de etanol no interior da célula da levedura. O modelo mostra os principais compostos químicos e substratos das vias propostas, bem como as enzimas responsáveis pela conversão/utilização dos compostos



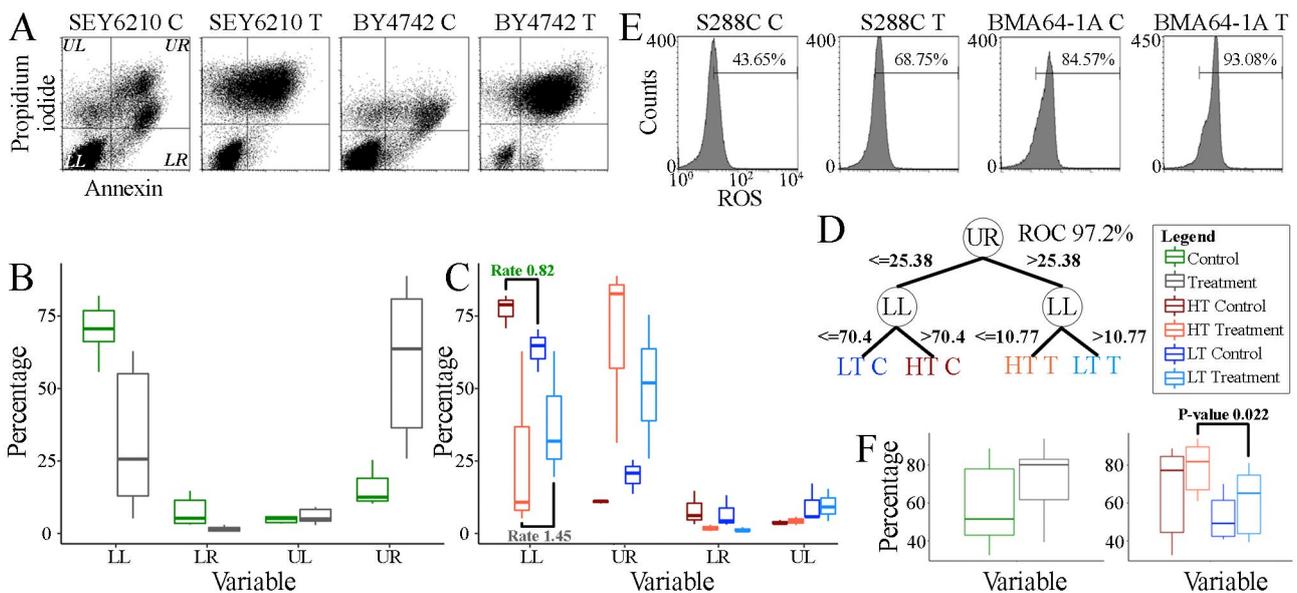
**Figura 16:** Heatmap dos genes diferencialmente expressos na provável via de tamponamento de etanol.

#### 4.5. Citometria de fluxo e ensaio SDH

Uma vez que os genes de resposta ao estresse oxidativo, condição em que há um desequilíbrio entre a produção e remoção de espécies reativas de oxigênio (ROS), foram sobre expressos em ambos os grupos de linhagens, um ensaio de ROS foi realizado (**Tabela 8** e **Figura 17F**). O resultado reportou um aumento na quantidade de células acumulando ROS significativamente durante um estresse máximo de etanol a uma taxa similar, tanto para HT como para LT, sendo 1,23 e 1,15, respectivamente. Apesar das divergências entre os grupos tratamento vs controle não serem grandes, comparando a acumulação de ROS entre HT vs LT na situação de tratamento, as diferenças são estatisticamente significativas (valor de p-value= 0,02246), enquanto que o acúmulo nas células na situação controle não divergem (valor de p-value = 0,248). Além disso, a taxa de LT vs HT em células tratadas é de 0,76, o que significa que as linhagens HTs têm mais células acumulando ROS do que as linhagens LTs (**Tabela 8**).

**Tabela 8:** Comparação dos dados de ensaios ROS, H2AX e SDH. **Avg C:** Média no controle; **Avg T:** Média no Tratamento; **FC (T/C):** Fold-change (Tratamento/Controle).

Grupos	Dados ROS			Dados H2AX			Dados SDH		
	Avg C	Avg T	FC (T/C)	Avg C	Avg T	FC (T/C)	Avg C	Avg T	FC (T/C)
HT	64.45	79.43	1.23	10.02	19.55	1.95	3.65	4.75	1.30
LT	52.52	60.39	1.15	8.46	21.57	2.55	2.88	3.25	1.13
Rate (LT/HT)	0.81	0.76	-	0.84	1.10	-	0.79	0.68	-



**Figura 17:** Ensaio de viabilidade celular e contagem de células que acumulam ROS. **A:** exemplos de ensaio de citometria de fluxo para medir a viabilidade celular. É mostrado uma linhagem LT (SEY6210) e uma linhagem HT (BY4742) na condição de controle (C) e tratamento (T). O eixo é relatado no primeiro *box*; **B:** análise de viabilidade celular para todas as linhagens sob controle e tratamento; **C:** análise de viabilidade celular em HTs e LTs. As "taxas" são baseadas nas média das medições de LT divididas pela média das medições de HT; **D:** árvore de decisão para modelar o relacionamento entre variáveis LL e UR. Os números nas arestas são a porcentagem de células sob a condição da descrição do nó (UR ou LL); **E:** exemplos de ensaio de citometria de fluxo para medir o número de células que acumulam ROS. É apresentada uma linhagem LT (S288C) e outra linhagem HT (BMA64-1A) sob controle (C) e tratamento (T); **F:** análise de ROS para todas as linhagens sob controle e tratamento (*boxplot* esquerdo) e dentro de HTs e LTs (*boxplot* direito). O valor de *p* foi calculado usando o teste T (pareado); LL significa inferior esquerdo (células viáveis), LR significa inferior direito (células sob apoptose inicial), UL significa superior esquerdo (células mortas) e UR significa superior direito (morte avançada); A legenda na caixa representa os dados em *boxplots* e na árvore de decisão.

Uma vez que a principal fonte de ROS intracelular é a mitocôndria (MARIS et al., 2001), foi realizado um ensaio de succinato desidrogenase (SDH) e a ANOVA de fator único, considerando os valores de triplicatas experimentais para cada linhagem de levedura e condição experimental (tratamento ou controle). O resultado reportou

diferenças significativas na quantidade de mitocôndrias ou na atividade da SDH após o tratamento para as linhagens BMA64-1A, BY4742 e quando se compara os resultados de todas as HTs (**Tabela 9**). Uma análise das taxas dos valores aferidos reportou que as linhagens HTs têm SDH mais ativo (ou um maior número de mitocôndrias) do que as linhagens LTs (**Tabela 8**).

**Tabela 9:** Resultados dos testes estatísticos do ensaio de SDH. \*: Diferenças estatisticamente significativas (p-value <0,05).

<b>Linhagens</b>	<b>p-value</b>
BMA64-1A Control x BMA64-1A Treatment	<b>0.002*</b>
BY4742 Control x BY4742 Treatment	<b>0.001*</b>
X2180-1A Control x X2180-1A Treatment	0.205
BY4741 Control x BY4741 Treatment	0.496
S288c Control x S288c Treatment	0.097
SEY6210 Control x SEY6210 Treatment	0.718
HT Control x HT Treatment	<b>0.025*</b>
LT Control x LT Treatment	0.135

Como os ROS podem ser uma fonte de dano ao DNA (MANI; CHINNAIYAN, 2010), um ensaio de citometria de fluxo foi realizado utilizando o anticorpo H2AX para medir a porcentagem de células com dano significativo no DNA, tanto na situação controle como no tratamento. Os resultados reportam um pequeno número de células sofrendo danos no DNA em todas as linhagens (**Tabela 10**). No entanto, o tratamento aumenta o dano ao DNA em todas as linhagens em comparação com a condição controle (*fold-change* de 1,95 e 2,55 para HTs e LTs, respectivamente) (**Tabela 8**). Assim, os resultados demonstram que as LTs sofrem mais danos do que as linhagens HT.

**Tabela 10:** Porcentagem de células acumulando danos ao DNA (ensaio utilizando o anticorpo H2AX).

<b>Linhagem</b>	<b>Controle</b>	<b>Tratamento</b>	<b>Fold-change (tratamento/controle)</b>
BMA64-1A	18.03	22.09	1.23
BY4742	7.88	18.15	2.30
X2180-1A	4.14	18.42	4.45
BY4741	12.56	23.72	1.89
S288C	2.88	6.01	2.09
SEY6210	9.93	34.98	3.52

Finalmente, o ensaio de viabilidade celular (FC anexina) foi realizado em colaboração com ALMEIDA (ALMEIDA, 2017) para compreender a sobrevivência celular no estresse por etanol (**Figura 17A**). O teste de independência do Chi-quadrado, considerando as porcentagens de células em cada quadrante (LL, LR, UL e UR), reportou que em quase todas as situações as diferenças entre tratamento e controle são significativas (**Tabela 11**). Ainda, reporta-se um maior número de células viáveis no controle em relação ao tratamento, ao passo que o tratamento induziu um maior número de células à morte avançada; também observa-se que o número de células em início de apoptose e de células mortas não variam muito após o tratamento (**Figura 17B**). Além disso, as taxas calculadas comparando as células viáveis em LTs vs HTs, reportam que as HTs estão mais propensas a sofrerem com o estresse etanólico (**Figura 17C**); a árvore de decisão confirma a provável maior resistência das LTs (**Figura 17D**).

**Tabela 11:** Diferença entre os quadrantes do ensaio de viabilidade celular. \*: Diferenças estatisticamente significativas (valor de  $p < 0,05$ ).

<b>Linhagens</b>	<b>p-value</b>
BMA64-1A Control x BMA64-1A Treatment	<b>3.906e-13*</b>
BY4742 Control x BY4742 Treatment	<b>2.2e-16*</b>
X2180-1A Control x X2180-1A Treatment	<b>2.2e-16*</b>
BY4741 Control x BY4741 Treatment	0.8554
S288c Control x S288c Treatment	<b>0.001541*</b>
SEY6210 Control x SEY6210 Treatment	<b>0.000000263*</b>
HT Control x HT Treatment	<b>2.2e-16*</b>
LT Control x LT Treatment	<b>0.01955*</b>

#### 4.6. Rede global integrada e análises estatísticas

Os procedimentos aplicados para a obtenção da rede integrada permitiram gerar 16 redes diferentes (6 linhagens X 2 condições, mais 2 grupos X dois estados) (**Tabela 12**).

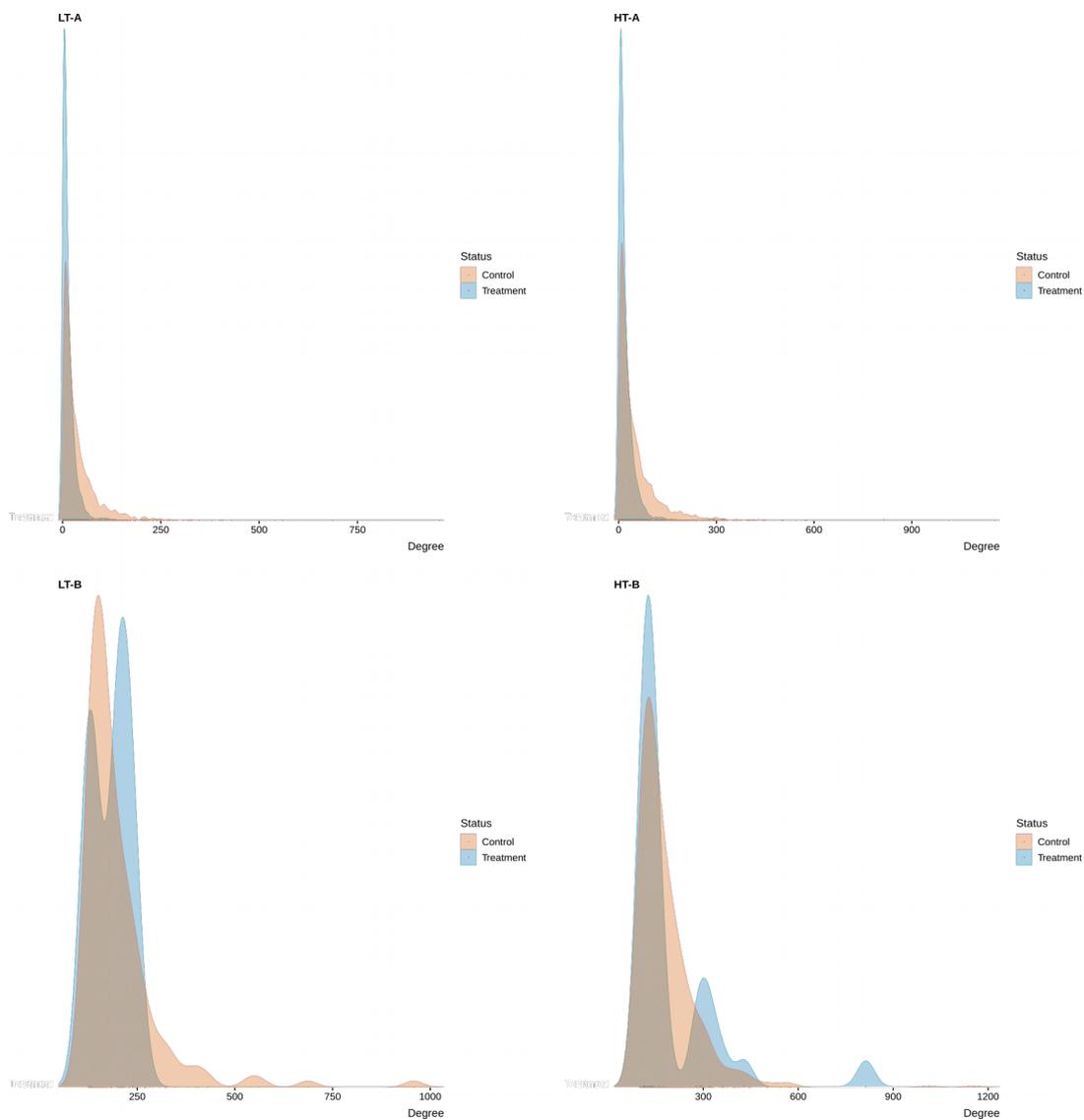
**Tabela 12:** Estatísticas gerais das redes específicas. **C:** controle; **T:** tratamento.

Strain_Graph	Density	Transitivity	Diameter	Average_Path_Length	Assortativity_Degree	Number_of_Edges	Number_of_Vertex
BMA64-1A_C	0.013	0.189	5	2.421	-0.087	150306	4808
X2180-1A_C	0.018	0.198	5	2.360	-0.086	107316	3415
BY4742_C	0.017	0.186	5	2.374	-0.092	104525	3498
S288C_C	0.015	0.211	5	2.435	-0.079	143186	4319
SEY6210_C	0.016	0.188	5	2.351	-0.087	130385	3926
BY4741_C	0.013	0.133	5	2.362	-0.125	78877	3432
BMA64-1A_T	0.011	0.156	6	2.407	-0.109	129636	4695
X2180-1A_T	0.008	0.115	6	2.711	-0.097	40816	3116
BY4742_T	0.008	0.119	6	2.759	-0.082	38775	3067
S288C_T	0.006	0.102	7	2.792	-0.100	44446	3672
SEY6210_T	0.009	0.114	5	2.619	-0.096	56823	3501
BY4741_T	0.012	0.154	6	2.515	-0.102	65086	3192
HT_control	0.018	0.196	5	2.392	-0.080	79133	2943
HT_treated	0.008	0.123	7	2.826	-0.080	24713	2448
LT_control	0.018	0.176	5	2.395	-0.114	45609	2209
LT_treated	0.008	0.133	7	3.145	-0.091	11583	1682

Para as diferentes redes geradas, também foram obtidas 13 medidas topológicas para cada nó da rede. O teste de Shapiro-Wilk revelou que os dados não seguem uma distribuição normal, sendo assim, o teste estatístico de Kruskal-Wallis foi aplicado para comparar se havia diferença significativa nas topologias das redes de tratamento em comparação com as redes de controle. As diferenças estão descritas na **Tabela 13**. Interessantemente, os resultados mostraram que as diferenças entre tratamento vs controle, principalmente quanto ao *degree* (K) e *eigenvector centrality* (EIG), são estatisticamente significantes (**Tabela 13**) e que a cauda da distribuição de *degree* na rede LT tem uma finalização precoce quando comparada com a rede no tratamento, evidenciando uma perda de *hubs* mais acentuada nesse fenótipo (**Figura 18**).

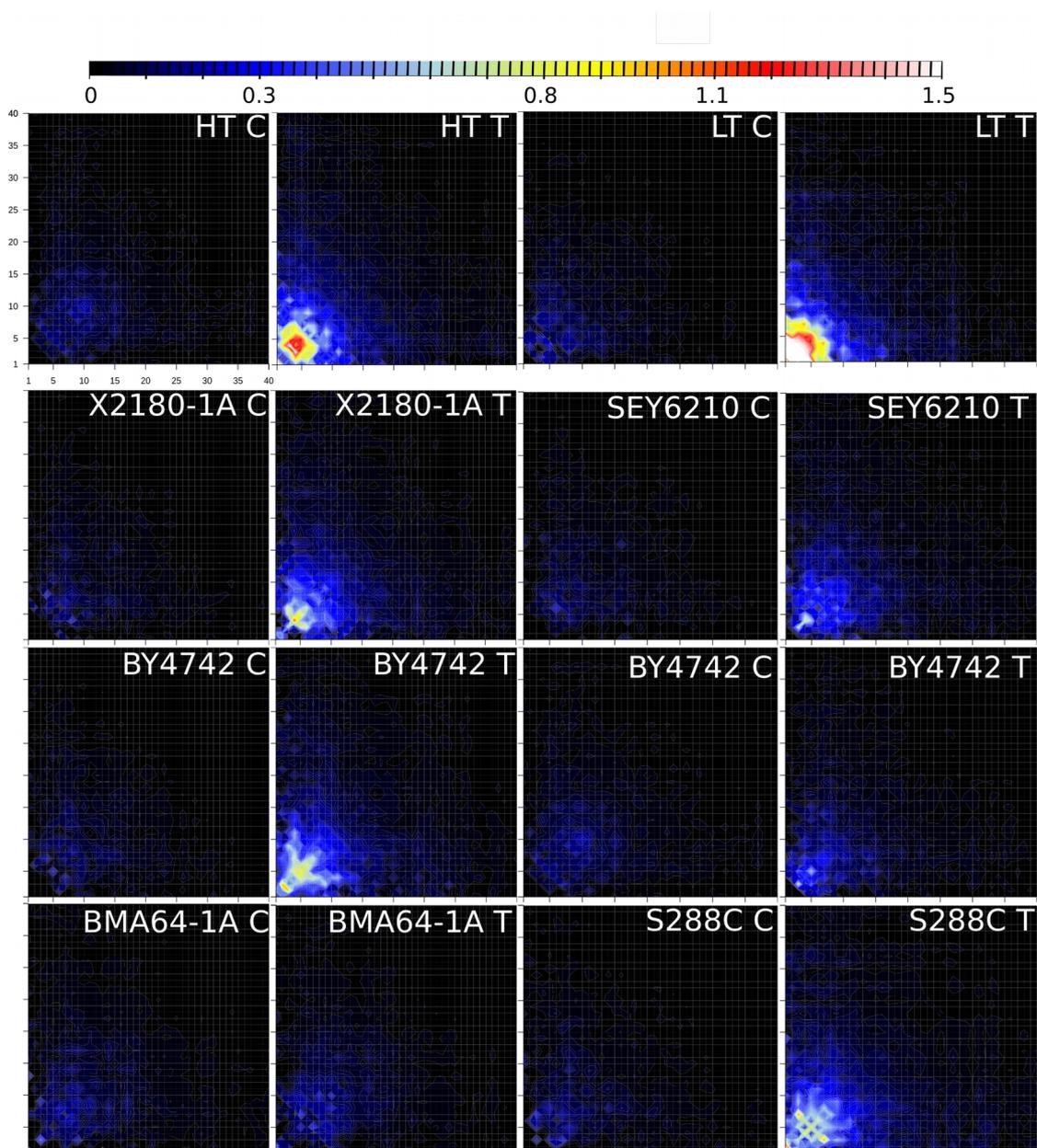
**Tabela 13:** Diferença estatística em medidas topológicas dos nós. **ECC:** eccentricity; **K:** degree; **EIG:** eigenvectors; **BET:** betweenness; **CLOS:** closeness; **PR:** page rank; **S:** strength; **HS:** hub score; **C:** coreness; **SC:** subgraph centrality; **BC:** burtconstraint; **T:** transitivity; **K<sub>nn</sub>:** Average nearest neighbor degree; **+**: aumento significativo no tratamento com relação ao controle; **-**: diminuição significativa no tratamento com relação ao controle; **NS:** Não significativo.

Strain	ECC	K	EIG	BET	CLOS	PR	S	HS	C	SC	BC	T	K <sub>nn</sub>
LT	+	-	-	+	-	+	-	-	-	-	+	-	-
HT	+	-	-	+	-	+	-	-	-	-	+	-	-
BMA64-1A	+	-	-	-	+	NS	-	-	-	-	+	+	NS
X2180-1A	+	-	-	+	-	+	-	-	-	-	+	-	-
BY4742	+	-	-	+	-	+	-	-	-	-	+	-	-
S288C	+	-	-	+	-	+	-	-	-	-	+	-	-
BY4741	+	-	NS	+	+	+	-	NS	-	-	+	-	-
SEY6210	+	-	-	+	-	+	-	-	-	-	+	-	-



**Figura 18:** Distribuição dos *degrees* para a rede controle e rede tratamento. **LT-A:** distribuição para toda a faixa de *degrees* para a rede integrada LT; **LT-B:** detalhe da cauda de LT-A; **HT-A:** distribuição para toda a faixa de *degrees* para a rede integrada HT; **HT-B:** detalhe da cauda de HT-A.

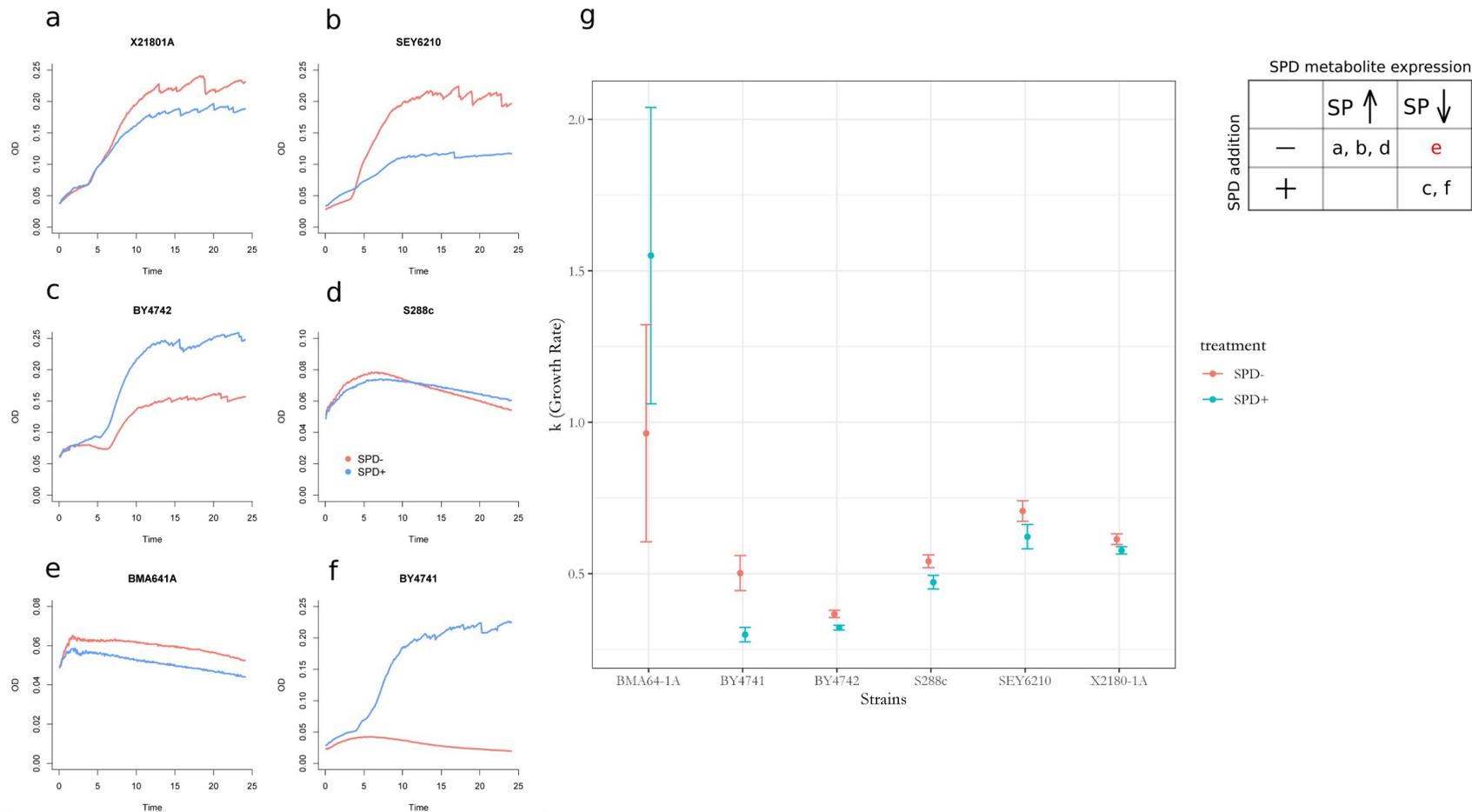
Ademais, as redes de tratamento apresentam maiores probabilidades de conexão entre os nós de menor grau de acordo com a análise normalizada da função DDC (**Figura 19**). Isso fornece uma evidência extra de que hubs estão perdendo conexões, muito provavelmente mantendo somente as conexões entre si em um estado basal de regulação, uma vez que no tratamento seu *degree* é menor (**Figura 19**).



**Figura 19:** Análise da função DDC normalizada para redes de controle e tratamento. A escala mostra a probabilidade de que dois genes com graus  $k$  e  $k'$  interajam. Os eixos  $x$  e  $y$  representam o *degree*. Cores mais escuras significam interação reduzida, enquanto cores brancas significam propensão para interação genética.

#### 4.7. Experimentos de crescimento celular acrescidos de espermidina

Os dados de crescimento populacional em um meio com estresse máximo de etanol (v/v máximo para cada linhagem), suplementados ou não com espermidina, demonstraram que a adição de espermidina aumentou a população máxima de BY4742 e BY4741, enquanto que nas outras linhagens a adição de espermidina inibiu o crescimento em comparação com o grupo de controle. Curiosamente, nas duas linhagens citadas, o metabólito espermidina está sub abundante (**Apêndice 2**) enquanto que nas linhagens SEY6210, S288C e X2180-1A, tal metabólito está sobre abundante. Assim, é possível afirmar que a adição de espermidina nas linhagens em que este metabólito é menos abundante após o tratamento leva ao aumento do crescimento quando comparado ao controle. Enquanto tal adição nas linhagens em que a espermidina já é muito abundante após o tratamento acarreta na inibição da taxa de crescimento. A única linhagem em que essa hipótese não pode ser aplicada é a BMA64-1A, a qual possui sub abundância de espermidina após o tratamento e que um meio enriquecido com tal metabólito não aumenta o crescimento celular (**Figura 20**).



**Figura 20:** Crescimento celular em meio com estresse máximo de etanol e enriquecido com espermidina. **a-f:** curvas de crescimento. **Eixo X:** representa o tem em horas. **Eixo Y:** OD 600<sub>nm</sub>. **Vermelho:** meio com estresse máximo de etanol sem espermidina (SPD-); **Azul:** meio com estresse máximo e acrescido de espermidida (SPD+); **g:** O boxplot representa os valores de k obtidos a partir do ajuste no modelo de Verhulst com intervalo de confiança de 95%. A matriz localizada no canto superior direito representa as observações feitas a partir do experimento em meio acrescido com espermidina, no qual “-” representa atividade inibitória do crescimento e “+” representa um aumento no crescimento. **SP↑:** representa a situação de sobre abundância de espermidina (**Apêndice 2**). **SP↓:** representa a situação de sub abundância de espermidina (**Apêndice 2**). As letras representam as linhagens dos gráficos a-g sendo o f em vermelho a única linhagem que não se ajustava ao observado.

#### 4.8. Rede global integrada e análises de Machine Learning

Nesta etapa foi efetuada a tentativa de aprendizado utilizando as matrizes de conexão de genes representadas de maneira binária e as matrizes de conexão com os dados de expressão discretizados. Foi possível observar que o processo de aprendizado de máquina os dados de expressão discretizados, gerou uma performance preditiva superior com um ROC >0,8 para alguns tipos de redes (**Tabela 14**).

**Tabela 14:** Performance preditiva em diferentes matrizes. **4 classes:** as classes utilizadas foram "control\_LT", "treatment\_LT", "control\_HT", "treatment\_HT", ou seja, a condição e o fenótipo; **2 classes:** as classes utilizadas foram "treatmet\_LT", "treatment\_HT", ou seja, somente o fenótipo; **bin:** matriz baseada em dados binários de conexões; **exp:** baseada nas relações de expressões discretizadas entre os nós. **\*\*:** melhores performances preditivas.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Matrix
*	*	*	*	*	*	*	*	4 classes bin Integrada
0,692	0,308	0,743	0,692	0,675	0,432	0,798	0,808	2 classes bin Integrada
0,622	0,378	0,734	0,622	0,57	0,337	0,654	0,662	2 classes bin Metabólica
0,618	0,382	0,696	0,618	0,576	0,304	0,694	0,711	2 classes bin Regulatória
0,681	0,319	0,726	0,681	0,665	0,404	0,783	0,794	2 classes bin PPI
0,602	0,398	0,776	0,602	0,527	0,335	0,596	0,602	2 classes bin lncRNA
0,49	0,17	0,619	0,49	0,464	0,376	0,753	0,544	4 classes bin Metabólica
0,38	0,207	0,425	0,38	0,341	0,199	0,67	0,406	4 classes bin Regulatória
*	*	*	*	*	*	*	*	4 classes bin PPI
0,32	0,227	0,477	0,32	0,292	0,145	0,589	0,347	4 classes bin lncRNA
0,731	0,269	0,731	0,731	0,731	0,462	**0,856	0,855	2 classes exp Metabólica
0,721	0,279	0,732	0,721	0,718	0,453	**0,833	0,843	2 classes exp Regulatória
0,78	0,22	0,79	0,78	0,778	0,57	**0,891	0,892	2 classes exp PPI
0,68	0,32	0,802	0,68	0,643	0,466	0,673	0,669	2 classes exp lncRNA
0,21	0,263	0,21	0,21	0,208	-0,053	0,7	0,342	4 classes exp Metabólica
0,265	0,245	0,265	0,265	0,262	0,02	0,709	0,367	4 classes exp Regulatória
*	*	*	*	*	*	*	*	4 classes exp PPI
0,305	0,232	0,307	0,305	0,273	0,074	0,609	0,309	4 classes exp lncRNA

## 5. Discussão

### 5.1. Remodelando o diauxic shift: mecanismo fundamental para a tolerância ao etanol

As análises de expressão diferencial mapeando os dados de RNA-Seq nos genomas específicos de cada linhagem permitiu identificar que cada genoma responde de forma específica ao estresse por etanol (**Figura 12**). Além disso, pela primeira vez foi observado que as diferenças na razão de proporção de genes diferencialmente expressos entre os grupos HT e LT são significativas (**Tabela 4**).

Apesar das peculiaridades de cada linhagem quanto à expressão gênica durante o estresse máximo ao etanol, foi observado um número de genes diferencialmente expressos comuns entre as linhagens de cada grupo (HT ou LT) (**Figura 12**). Desse modo, fornecemos evidências adicionais para um mecanismo de resposta ao estresse comum entre as diferentes linhagens (MA; LIU, 2010; STANLEY et al., 2010a).

Comparando os dados aqui obtidos com os dados dos genes comumente expressos em situações de estresse fermentativo segundo Auesukaree (AUESUKAREE, 2017) (**Tabela 5**), pode-se citar que o estresse osmótico parece não ser um fator determinante no efeito fenotípico de tolerância ao etanol. No entanto, os genes relacionados à resposta ao calor, oxidação e etanol, estão sobre expressos principalmente em linhagens HT. Contudo, parece que esses genes estão atuando na produção de agentes anti-oxidantes (genes SOD), na manutenção de proteínas (HSPs) e da membrana (gene GSH1) principalmente em linhagens HT. Isto parece reforçar o fato de que o tratamento de espécies reativas de oxigênio parece ser mais importante do que a acidez celular (CHAROENBHAKDI et al., 2016; AUESUKAREE, 2017), uma vez que o acúmulo de ROS pode levar a vias mitocondriais de apoptose (PÉREZ-GALLARDO et al., 2013).

Genes sobre expressos comuns nos grupos HT e LT exibiram termos GO relacionados a metabolismo, energia, mitocôndrias e respostas a elementos com oxigênio, enquanto que genes sub expressos estavam relacionados a processamento de RNAs e ribossomos. Este cenário é compatível com alguns resultados já descritos na literatura para algumas linhagens (NAVARRO-TAPIA et al., 2016; ZYRINA et al., 2017). Isso demonstra que apesar da divergência entre linhagens referente aos genes diferencialmente expressos, existe sim uma coerência nos processos biológicos alterados na célula da levedura (**Figura 13B**).

Os genes relacionados à termos GO de regulação da estabilidade proteica e dobramento de proteínas, estavam sobre-expressos em linhagens HT e não eram listados em linhagens LT (**Tabela 6, Apêndice 1**). A estabilidade de proteínas é crucial na manutenção das membranas biológicas, sendo então um fator importante na tolerância ao etanol (NARAYANAN et al., 2016; ISHMAYANA; KENNEDY; LEARMONTH, 2017). Isto chamou a atenção uma vez que aqui reporta-se pela primeira vez o termo “formação de eisossomos” em estudos sobre etanol.

Eisossomos são invaginações que podem aumentar a tensão da membrana plasmática (KABECHE; HOWARD; MOSELEY, 2015) e sugerimos que em um cenário de maior concentração de etanol, os eisossomos poderiam contribuir para a desacidificação da célula, pois além de ajudar na estabilidade da membrana plasmática estes também podem estar envolvidos no fluxo de prótons (DOUGLAS et al., 2011). Charoenbhakdi *et al* (CHAROENBHAKDI et al., 2016) avaliaram uma linhagem BY4742 mutante para os genes PMA1 e PMA2 (genes que fazem parte do complexo da bomba de prótons V-ATPase) manteve seu pH estável. De fato, nos nossos resultados, os genes PMA1 ou PMA2 estavam geralmente sub expressos em todas as linhagens onde foram significativos e, dessa forma, os eisossomos podem constituir um sistema auxiliar para a remoção de

prótons do interior da célula, o que poderia explicar por que a linhagem BY4742 aqui utilizada (classificada como HT) manteve seu pH estável.

O termo relacionado ao metabolismo de carnitina foi representado entre os GOs observados na linhagem mais tolerante, a BMA64-1A (**Apêndice 1**). A carnitina tem um papel na resposta a estresse oxidativo em *S. cerevisiae* e em outros organismos (FRANKEN et al., 2008; AUGUSTYNIAK; SKRZYDLEWSKA, 2009). Desse modo a carnitina poderia contribuir para a diminuição da acidez e de ROS intracelulares, permitindo que, mesmo em concentrações maiores de etanol, a linhagem BMA64-1A resista melhor a perda de homeostase.

O enriquecimento de vias do KEGG revelou que a via do ciclo do citrato estava sobre representada. Considerando as expressões dos transportadores de HTX1 e 2 a glicose provavelmente não constituía a principal fonte de carbono para o ciclo do ácido tricarboxílico. Assim, a expressão das enzimas ALD indicam a atividade mitocondrial de utilização do álcool como fonte de carbono no ciclo do citrato, como já previamente reportado (CHANDLER et al., 2004). Para isso, as enzimas ADH podem estar atuando na conversão do álcool em acetaldeído (PISKUR et al., 2006) e as ALDs podem estar transformando o acetaldeído em acetado para, enfim, ser convertido em Acetil-CoA pela ACS1, finalizando com a entrada desse composto no ciclo do citrato. Esse fato pode indicar que a tolerância superior da linhagem BMA64-1A pode estar relacionada as quatro diferentes ADHs sobre expressas em seu transcriptoma, enquanto que nas demais linhagens somente duas ADHs estão sobre-expressas. Reforçando nosso modelo, o ensaio SDH da mitocôndria permitiu concluir que a atividade enzimática mitocondrial é superior em linhagens HT. De fato, há uma maior correlação da expressão de genes mitocondriais com genes de resposta a estresse em *S. cerevisiae*, correlação essa que

está ausente em outras espécies de fungos (PISKUR et al., 2006), reforçando a importância da mitocôndria na resposta ao estresse por etanol.

A utilização do etanol como substrato para o ciclo do ácido cítrico é conhecida como *diauxic shift* (PISKUR et al., 2006), e já foi descrito como parte do metabolismo do piruvato (ver (PRONK; YDE STEENSMA; VAN DIJKEN, 1996)). No entanto, este é o primeiro relato que relaciona o *diauxic shift* à tolerância ao etanol. Ademais, como esta via ocorre principalmente na mitocôndria, ela pode ser responsável pelo aumento de ROS intracelulares na levedura (MARIS et al., 2001). De fato, aqui demonstramos que as linhagens HT possuem maior atividade mitocondrial e mais ROS que linhagens LT (**Tabela 8**) e isto está de acordo com o modelo aqui proposto (**Figura 15**).

Por fim, os demais mapas significativamente afetados estão relacionados a síntese de ribossomos, produção e transporte de RNA e aminoácidos, indicando uma pausa de metabolismos centrais, fato esse já observado com relação ao etanol por Chandler *et al* (CHANDLER et al., 2004). Apesar de ainda não estar bem esclarecida, há uma relação entre os danos ao DNA e à tradução (BEGLEY et al., 2007). Assim, estes mapas podem estar sendo reportados como afetados pelo tratamento principalmente por conta dos danos ao DNA causados pelo estresse (**Tabela 8**).

### 5.2. A espermidina tem um papel importante para a tolerância ao etanol

A espermidina é de especial interesse neste ponto, pois exerce influência nos processos de crescimento celular (BALASUNDARAM; TABOR; TABOR, 1991). Sabe-se também que a espermidina desempenha um papel importante no aumento da expectativa de vida de vários tipos de células (EISENBERG et al., 2009) e apresentou um aumento na

tolerância aos inibidores da fermentação refletindo em uma maior produtividade do etanol (KIM et al., 2017).

Os dados metabolômicos mostraram que o estresse por etanol afeta a produção de espermidina de forma diferente para cada linhagem (**Figura 20**). Para as linhagens nas quais a espermidina é menos abundante durante o estresse máximo, após o aumento da concentração deste composto no meio, houve um aumento no crescimento populacional; enquanto que em linhagens com maior abundância de espermidina o acréscimo levou a uma redução da taxa de crescimento (**Figura 20**), confirmado com maior população máxima recuperada do modelo (dados não mostrados). A única exceção nesse modelo foi a linhagem BMA64-1A. Desse modo, a espermidina parece ter um papel na tolerância ao estresse e no crescimento populacional, no entanto, estes efeitos dependem da sua concentração e abundância.

Dentre as vias em que podemos encontrar a espermidina estão a de transportadores de membrana e metabolismos de arginina e prolina (**Tabela 7**). Nesta última via, a prolina está mais abundante em ambos os grupos de linhagens (**Figura 14**), o que reforça o papel do metabolismo da prolina na sobrevivência das linhagens em condições de estresse ao etanol (NAVARRO-TAPIA et al., 2016; TAKAGI; TAGUCHI; KAINO, 2016). Já foi observado que a ausência do gene SPE1 (responsável pela biossíntese de poliamina e que reflete na diminuição da espermidina) diminui o tempo de duplicação das células, enquanto as células do tipo selvagem apresentam um maior tempo de duplicação (EISENBERG et al., 2009).

### 5.3. O estresse por etanol ocasiona um expressivo rewiring das redes: os módulos basais tendem a ser os que se mantêm mais ativos

Utilizando os atributos topológicos gerais das redes tratamento com relação aos atributos das redes controle (**Tabelas 12 e 13**), foi possível inferir a ativação de vias especializadas de resposta tardia. Essa inferência se baseia no aumento do diâmetro e do tamanho dos caminhos na rede (**Tabela 12**), uma vez que caminhos mais longos favorecem a modularidade da rede (XU et al., 2011) e também fornecem uma ideia do tempo de atraso (*delay*) entre a ativação de um regulador principal (por exemplo, um fator de transcrição ou uma kinase) e sua resposta regulatória na via (KLEIN et al., 2012). Assim, a célula está preparada para respostas rápidas e dinâmicas de acordo com os estímulos detectados no meio (MANGAN; ZASLAVER; ALON, 2003).

Além disso, genes com menor *degree* estão sendo desativados (provavelmente enzimas nos finais de cadeias regulatórias). Isso pode ser inferido pelo leve aumento da assortatividade, tal como reportado por Barrat et al. (BARRAT et al., 2004). Também, a diminuição da densidade (**Tabela 12**) indica uma diminuição do número de aresta/nó na rede (XU et al., 2011), mostrando uma perda de conexões internas. Assim, rotas alternativas para o fluxo da informação na rede podem estar menos disponíveis, refletindo em um aumento da *burst constraint* (BC) (BURT, 2004; BUSKENS; VAN DE RIJT, 2008). Logo, provavelmente temos cadeias regulatórias menos complexas após os tratamentos.

A diminuição da transitividade, que mostra o número de *loops*/cliques de tamanho 3 (WASSERMAN; FAUST, 1994), é um indício de que as vias de sinalização e regiões de interação genética negativa podem estar se desestruturando (IGNATIUS PANG; GOEL; WILKINS, 2018). Desta maneira, esses dados corroboram com a hipótese de que somente vias essenciais e vias especializadas de resposta podem estar ativas.

Ao observar as diferenças entre medidas topológicas (**Tabela 13**) é evidente que o degree ( $K$ ) está diminuindo de maneira significativa nos tratamentos. Ao observar a diferença entre a distribuição de  $K$  nas redes “tratamento” com relação as redes “controle”, o número de *hubs* intermediários da rede HT fica mais evidente (**Figura 18**).

O aumento de *hubs* intermediários pode ser inferido através da diminuição significativa da *eigenvector centrality* (EIG) (**Tabela 13**). Isso demonstra que os nós de menor importância, aqueles com menos conexões (menor  $k$ ) e provavelmente genes nos finais de cadeias regulatórias, estão perdendo conexões com nós de maior importância (provavelmente reguladores/kinases no topo de cadeias regulatórias) (LOHMANN et al., 2010); isso pode aumentar o número de pontes somente entre *hubs*, diminuindo assim o valor de EIG na rede (mantendo somente cadeias essenciais entre reguladores transcricionais).

Contudo, inferimos então que a situação de estresse parece estar exigindo que somente os módulos basais permaneçam ativos, sendo provavelmente os responsáveis por manter conexões essenciais no controle sobre outras conexões na rede (como exemplo destes módulos basais podemos citar as vias relacionadas ao metabolismo de carboidratos, cofatores, crescimento e energia (**Tabela 6**)). Essas características, por sua vez, são de nós com alto *betweenness* (BET) (DONCHEVA et al., 2012; KLEIN et al., 2012), o que justificaria o aumento significativo nos valores dessa medida de centralidade (**Tabela 13**).

#### 5.4. Rede global integrada e análises de Machine Learning

A análise da rede global integrada exibiu performances inferiores a do estudo detalhado nas Figuras 2 e 3 do **Apêndice 3**. Por conta destes resultados, a performance das redes foi avaliada de maneira individual. Ao alterar a estrutura da matriz e utilizar os

dados de expressão, é possível perceber a melhora da performance preditiva, principalmente quando se utilizam duas classes (HT e LT) (**Tabela 14**).

Também é visível que a performance preditiva é melhor em redes PPI, seguidas das redes metabólicas, das redes regulatórias e por fim das redes lncRNAs (**Tabela 14**). As interações representadas na rede PPI também podem constituir um conjunto mais relevante de interações, uma vez que redes PPI podem ter sido produzidas seguindo tendências experimentais (ROLLAND et al., 2014).

Desse modo, visto que o *software* Weka já efetua a validação cruzada particionando o conjunto de dados, e dada a melhora da performance somente alternando a maneira com que os dados são representados na matriz, podemos levantar a hipótese de que a quantidade de informação pode também estar gerando um *overfitting* do modelo. Essa hipótese é reforçada pelo grande número de folhas e pelo tamanho das árvores (504 e 1007, respectivamente, somente para a rede PPI na melhor situação testada). Assim, é possível concluir que aplicar métodos de seleção de atributos em conjunto com alterações na maneira com que os dados estão sendo representados, podem permitir o aumento da performance para algo próximo ao que foi reportado no **Apêndice 3**.

### 5.5. Visão geral da tolerância ao etanol em *S. cerevisiae*

De modo geral, as linhagens HT e LT utilizam estratégias diferentes frente ao estresse ao etanol. Inicialmente, um sistema de resposta comum é ativado, parte do metabolismo e das vias de síntese de RNA e ribossomos entram em pausa. Devido as mudanças na estrutura das redes metabólicas, regulatórias, interação proteína-proteína e lncRNAs-proteínas de cada linhagem, mecanismos de resposta ao estresse por etanol distintos são ativados em cada grupo.

As linhagens LT mostram uma tendência a manter a homeostase da célula com o ciclo celular ativo, ao custo de que elementos transponíveis ficam ativos e agentes anti oxidantes importantes podem estar sendo produzidos em menor proporção, evidenciados pela redução de mitocôndrias ou atividade mitocôndrial. Desse modo seu DNA sofre mais danos, o que pode contribuir para mudanças em seus mecanismos de regulação. Isto eventualmente poderia gerar mutações que levem a maior resistência no ambiente em que se encontra.

Já as linhagens HT tendem a tentar interromper seu ciclo celular e boa parte do metabolismo. Porém, elas mantêm a produção e reparo de proteínas para estabilizar as membranas celulares. A succinato desidrogenase está sobre expressa em linhagens HT, e o resultado do ensaio de atividade dessa enzima mostrou que aparentemente estas linhagens, após o tratamento, podem ter maior atividade deste complexo enzimático ou mais mitocôndrias que as linhagens LTs. Isso, em conjunto com a possível atividade de fluxo de prótons nos eisosomos, parece mostrar um mecanismo voltado para resistir melhor a entrada de íons  $H^+$  na célula para manter o pH intracelular o menos acidificado possível, enquanto buscam manter as membranas intactas. Assim, uma vez que o EtOH parece estar sendo utilizado para obtenção de carbono para o ciclo do ácido cítrico, a metabolização do álcool em conjunto ao controle de ROS podem permitir que as linhagens HT resistam a concentrações maiores de etanol.

## 6. Conclusões

Esta tese reporta a utilização de diferentes “ômicas”, abordagens de redes e inteligência artificial para investigar a tolerância ao etanol em *S. cerevisiae*. Além disso, é o primeiro relato em que “triplicatas fenotípicas” (3 linhagens diferentes, porém com o mesmo fenótipo) foram testadas sob o contexto de tolerância ao etanol utilizando as mesmas condições experimentais, protocolos e análises. Encontramos o conjunto mais rico de processos biológicos, funções moleculares e componentes celulares de genes diferencialmente expressos em comparação com trabalhos prévios.

Foram observados em média mais de 3900 genes diferencialmente expressos nas diferentes linhagens. Estes genes fornecem evidências de que o metabolismo e as vias de síntese de RNA e ribossomos entram em pausa. Além disso, alguns genes fazem parte dos sistemas de resposta a estresse oxidativo, por temperatura e álcool. Adicionalmente, a estabilidade das membranas celulares e o provável direcionamento do EtOH para a mitocôndria podem ser importantes para o fenótipo tolerante.

Os dados e conclusões aqui reportados são os primeiros a conseguir este nível de detalhes. Deste modo, este trabalho gera conhecimento que poderá ser aproveitado para enriquecer as discussões futuras sobre o fenômeno estudado e o desenvolvimento biotecnológico.

Finalmente, as evidências aqui apresentadas apontam pela primeira vez para a via de *diauxic shift* atuando como um *sistema de tamponamento do etanol*. Este sistema parece estar em funcionamento de maneira mais eficiente nas linhagens HT, caracterizando assim uma *possível assinatura sistêmica* a contribuir com tolerância ao etanol em *S. cerevisiae*.

## 7. Referências

- ALEXANDRE, H. et al. Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*. **FEBS Letters**, v. 498, n. 1, p. 98–103, jun. 2001.
- ALMEIDA, L. F. de. **Análise das linhagens de *Saccharomyces cerevisiae* expostas ao estresse por etanol**. 2017. Universidade Estadual Paulista “Júlio de Mesquita Filho”, Instituto de Biociências de Botucatu, 2017. Disponível em: <<http://hdl.handle.net/11449/152537>>.
- ALPER, H. et al. Engineering yeast transcription machinery for improved ethanol tolerance and production. **Science (New York, N.Y.)**, v. 314, n. 5805, p. 1565–1568, 8 dez. 2006.
- ASLANKOOHI, E. et al. Dynamics of the *Saccharomyces cerevisiae* Transcriptome during Bread Dough Fermentation. **Applied and Environmental Microbiology**, v. 79, n. 23, p. 7325–7333, 1 dez. 2013. Disponível em: <<http://aem.asm.org/cgi/doi/10.1128/AEM.02649-13>>.
- AUESUKAREE, C. Molecular mechanisms of the yeast adaptive response and tolerance to stresses encountered during ethanol fermentation. **Journal of Bioscience and Bioengineering**, v. 124, n. 2, p. 133–142, ago. 2017. Disponível em: <<http://dx.doi.org/10.1016/j.jbiosc.2017.03.009>>.
- AUGUSTYNIAK, A.; SKRZYDLEWSKA, E. I-Carnitine in the lipid and protein protection against ethanol-induced oxidative stress. **Alcohol**, v. 43, n. 3, p. 217–223, maio 2009. Disponível em: <<http://dx.doi.org/10.1016/j.alcohol.2008.12.005>>.
- BALASUNDARAM, D.; TABOR, C. W.; TABOR, H. Spermidine or spermine is essential for the aerobic growth of *Saccharomyces cerevisiae*. **Proceedings of the National Academy of Sciences**, v. 88, n. 13, p. 5872–5876, 1 jul. 1991. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.88.13.5872>>.
- BANKEVICH, A. et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, maio 2012. Disponível em: <<http://online.liebertpub.com/doi/abs/10.1089/cmb.2012.0021>>.
- BARRAT, A. et al. The architecture of complex weighted networks. **Proceedings of the National Academy of Sciences**, v. 101, n. 11, p. 3747–3752, 16 mar. 2004. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.0400087101>>.
- BEGLEY, U. et al. Trm9-Catalyzed tRNA Modifications Link Translation to the DNA Damage Response. **Molecular Cell**, v. 28, n. 5, p. 860–870, dez. 2007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1097276507006338>>.
- BLOHM, P. et al. Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. **Nucleic Acids**

- Research**, v. 42, n. D1, p. D396–D400, jan. 2014. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt1079>>.
- BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics (Oxford, England)**, v. 30, n. 15, p. 2114–20, 1 ago. 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24695404>>.
- BRUGGEMAN, F. J.; WESTERHOFF, H. V. The nature of systems biology. **Trends in Microbiology**, v. 15, n. 1, p. 45–50, 2007.
- BURT, R. S. Structural Holes and Good Ideas. **American Journal of Sociology**, v. 110, n. 2, p. 349–399, set. 2004. Disponível em: <<http://www.journals.uchicago.edu/doi/10.1086/421787>>.
- BUSKENS, V.; VAN DE RIJT, A. Dynamics of Networks if Everyone Strives for Structural Holes. **American Journal of Sociology**, v. 114, n. 2, p. 371–407, set. 2008. Disponível em: <<http://www.journals.uchicago.edu/doi/10.1086/590674>>.
- BUTLER, G. Fungal Sex and Pathogenesis. **Clinical Microbiology Reviews**, v. 23, n. 1, p. 140–159, 1 jan. 2010. Disponível em: <<http://cmr.asm.org/cgi/doi/10.1128/CMR.00053-09>>.
- CAMACHO, D. M. et al. Next-Generation Machine Learning for Biological Networks. **Cell**, v. 173, n. 7, p. 1581–1592, jun. 2018. Disponível em: <<https://doi.org/10.1016/j.cell.2018.05.015>>.
- CANTAREL, B. L. et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. **Genome Research**, v. 18, n. 1, p. 188–196, 2008.
- CARR, M.; BENSASSON, D.; BERGMAN, C. M. Evolutionary Genomics of Transposable Elements in *Saccharomyces cerevisiae*. **PLoS ONE**, v. 7, n. 11, p. e50978, 30 nov. 2012. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0050978>>.
- CHAKRABORTY, S. et al. Biomass to biofuel: a review on production technology. **Asia-Pacific Journal of Chemical Engineering**, v. 7, n. SUPPL. 3, p. S254–S262, ago. 2012. Disponível em: <<http://doi.wiley.com/10.1002/apj.1642>>.
- CHANDLER, M. et al. A genomic approach to defining the ethanol stress response in the yeast *Saccharomyces cerevisiae*. **Annals of Microbiology**, v. 54, n. 4, p. 427–454, 2004.
- 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, n. 10, p. 3121–3130, 15 maio 2016. Disponível em: <<http://aem.asm.org/lookup/doi/10.1128/AEM.00376-16>>.
- CHATR-ARYAMONTRI, A. et al. The BioGRID interaction database: 2015 update. **Nucleic acids research**, v. 43, n. Database issue, p. D470–8, jan. 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25428363>>.

- CHERRY, J. M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. **Nucleic acids research**, v. 40, n. Database issue, p. D700-5, 1 jan. 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22110037>>.
- CHONG, J.; XIA, J. MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. **Bioinformatics**, v. 34, n. 24, p. 4313–4314, 15 dez. 2018. Disponível em: <<https://academic.oup.com/bioinformatics/article/34/24/4313/5046255>>.
- COSTA, V. et al. Mitochondrial superoxide dismutase is essential for ethanol tolerance of Saccharomyces cerevisiae in the post-diauxic phase. **Microbiology**, v. 143, n. 5, p. 1649–1656, 1 maio 1997. Disponível em: <<http://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-143-5-1649>>.
- CREIXELL, P. et al. Pathway and network analysis of cancer genomes. **Nature Methods**, v. 12, n. 7, p. 615–621, 30 jun. 2015. Disponível em: <<http://www.nature.com/doi/10.1038/nmeth.3440>>.
- CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. **InterJournal**, v. Complex Sy, p. 1695, 2006. Disponível em: <<http://igraph.org>>.
- CUADROS-INOSTROZA, Á. et al. TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. **BMC Bioinformatics**, v. 10, n. 1, p. 428, 16 dez. 2009. Disponível em: <<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-10-428>>.
- DAS, S. et al. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. **International Journal of Computer Applications**, v. 115, n. 9, p. 31–41, 22 abr. 2015. Disponível em: <<http://research.ijcaonline.org/volume115/number9/pxc3902402.pdf>>.
- 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, n. 1, p. 89, jan. 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3698012&tool=pmcentrez&rendertype=abstract>>. Acesso em: 27 ago. 2014.
- DEMIRBAS, A. Tomorrow's biofuels: Goals and hopes. **Energy Sources, Part A: Recovery, Utilization, and Environmental Effects**, v. 39, n. 7, p. 673–679, 3 abr. 2017. Disponível em: <<http://dx.doi.org/10.1080/15567036.2016.1252815>>.
- DING, J. et al. **Tolerance and stress response to ethanol in the yeast Saccharomyces cerevisiae** *Applied Microbiology and Biotechnology*, nov. 2009. .
- DONCHEVA, N. T. et al. Topological analysis and interactive visualization of biological networks and protein structures. **Nature Protocols**, v. 7, n. 4, p. 670–685, 2012. Disponível em: <<http://dx.doi.org/10.1038/nprot.2012.004>>.

DOUGLAS, L. M. et al. Membrane Compartment Occupied by Can1 (MCC) and Eisosome Subdomains of the Fungal Plasma Membrane. **Membranes**, v. 1, n. 4, p. 394–411, 13 dez. 2011. Disponível em: <<http://www.mdpi.com/2077-0375/1/4/394>>.

EISENBERG, T. et al. Induction of autophagy by spermidine promotes longevity. **Nature Cell Biology**, v. 11, n. 11, p. 1305–1314, 4 nov. 2009. Disponível em: <<http://dx.doi.org/10.1038/ncb1975>>.

FABREGAT, A. et al. The Reactome Pathway Knowledgebase. **Nucleic Acids Research**, v. 44, n. D1, p. D481–D487, 14 jan. 2014. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt1102>>.

FAMILI, I. et al. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. **Proceedings of the National Academy of Sciences**, v. 100, n. 23, p. 13134–13139, 11 nov. 2003. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.2235812100>>.

FERNANDES, L. P. et al. Protein Networks Reveal Detection Bias and Species Consistency When Analysed by Information-Theoretic Methods. **PLoS ONE**, v. 5, n. 8, p. e12083, 18 ago. 2010. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0012083>>.

FRANKEN, J. et al. Carnitine and carnitine acetyltransferases in the yeast Saccharomyces cerevisiae: a role for carnitine in stress protection. **Current Genetics**, v. 53, n. 6, p. 347–360, 22 jun. 2008. Disponível em: <<http://link.springer.com/10.1007/s00294-008-0191-0>>.

FU, X. et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. **BMC genomics**, v. 10, p. 161, 2009.

GALLONE, B. et al. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. **Cell**, v. 166, n. 6, p. 1397–1410.e16, set. 2016. Disponível em: <<http://dx.doi.org/10.1016/j.cell.2016.08.020>>.

GE, H.; WALHOUT, a. J. M.; VIDAL, M. Integrating “omic” information: A bridge between genomics and systems biology. **Trends in Genetics**, v. 19, n. 10, p. 551–560, 2003.

GULLBERG, J. et al. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. **Analytical Biochemistry**, v. 331, n. 2, p. 283–295, ago. 2004. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0003269704003811>>.

GUPTA, A.; VERMA, J. P. Sustainable bio-ethanol production from agro-residues: A review. **Renewable and Sustainable Energy Reviews**, v. 41, p. 550–567, jan. 2015. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1364032114007084>>.

GUREVICH, A. et al. QCAST: quality assessment tool for genome assemblies. **Bioinformatics**, v. 29, n. 8, p. 1072–1075, 15 abr. 2013. Disponível em:

<<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btt086>>.

HALL, M. et al. The WEKA Data Mining Software : An Update. **SIGKDD Explorations**, v. 11, n. 1, p. 10–18, 2009.

HERRGARD, M. J. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. **Genome Research**, v. 16, n. 5, p. 627–635, 1 maio 2006. Disponível em: <<http://www.genome.org/cgi/doi/10.1101/gr.4083206>>.

HOFFMAN, D. E. et al. Changes in diurnal patterns within the *Populus* transcriptome and metabolome in response to photoperiod variation. **Plant, Cell & Environment**, v. 33, n. 8, p. no-no, 18 mar. 2010. Disponível em: <<http://doi.wiley.com/10.1111/j.1365-3040.2010.02148.x>>.

IGNATIUS PANG, C. N.; GOEL, A.; WILKINS, M. R. Investigating the Network Basis of Negative Genetic Interactions in *Saccharomyces cerevisiae* with Integrated Biological Networks and Triplet Motif Analysis. **Journal of Proteome Research**, v. 17, n. 3, p. 1014–1030, 2 mar. 2018. Disponível em: <<http://pubs.acs.org/doi/10.1021/acs.jproteome.7b00649>>.

ISHMAYANA, S.; KENNEDY, U. J.; LEARMONTH, R. P. Further investigation of relationships between membrane fluidity and ethanol tolerance in *Saccharomyces cerevisiae*. **World Journal of Microbiology and Biotechnology**, v. 33, n. 12, p. 218, 27 dez. 2017. Disponível em: <<http://dx.doi.org/10.1007/s11274-017-2380-9>>.

JACKMAN, S. D. et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. **Genome Research**, v. 27, n. 5, p. 768–777, maio 2017. Disponível em: <<http://genome.cshlp.org/lookup/doi/10.1101/gr.214346.116>>.

JIA, K.; ZHANG, Y.; LI, Y. Systematic engineering of microorganisms to improve alcohol tolerance. **Engineering in Life Sciences**, v. 10, n. 5, p. 422–429, out. 2010. Disponível em: <<http://doi.wiley.com/10.1002/elsc.201000076>>.

KABECHE, R.; HOWARD, L.; MOSELEY, J. B. Eisosomes provide membrane reservoirs for rapid expansion of the yeast plasma membrane. **Journal of Cell Science**, v. 128, n. 22, p. 4057–4062, 15 nov. 2015. Disponível em: <<http://jcs.biologists.org/cgi/doi/10.1242/jcs.176867>>.

KASAVI, C. et al. A system based network approach to ethanol tolerance in *Saccharomyces cerevisiae*. **BMC Systems Biology**, v. 8, n. 1, p. 90, 2014. Disponível em: <<http://www.biomedcentral.com/1752-0509/8/90>>.

KASAVI, C. et al. An integrative analysis of transcriptomic response of ethanol tolerant strains to ethanol in *Saccharomyces cerevisiae*. **Mol. BioSyst.**, v. 12, n. 2, p. 464–476, 2016. Disponível em: <<http://xlink.rsc.org/?DOI=C5MB00622H>>.

KIM, D.; LANGMEAD, B.; SALZBERG, S. L. HISAT: a fast spliced aligner with low memory requirements. **Nature Methods**, v. 12, n. 4, p. 357–360, 9 mar. 2015. Disponível em: <<http://www.nature.com/doi/10.1038/nmeth.3317>>.

KIM, S.-K. et al. Enhanced ethanol fermentation by engineered *Saccharomyces cerevisiae* strains with high spermidine contents. **Bioprocess and Biosystems Engineering**, v. 40, n. 5, p. 683–691, 24 maio 2017. Disponível em: <<http://link.springer.com/10.1007/s00449-016-1733-3>>.

KIM, S. et al. Evaluation and optimization of metabolome sample preparation methods for *Saccharomyces cerevisiae*. **Analytical Chemistry**, v. 85, n. 4, p. 2169–2176, 2013.

KIM, S. Y.; IMOTO, S.; MIYANO, S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. **Briefings in bioinformatics**, v. 4, n. 3, p. 228–35, 1 set. 2003. Disponível em: <<https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/4.3.228>>.

KLEIN, C. et al. Structural and dynamical analysis of biological networks. **Briefings in Functional Genomics**, v. 11, n. 6, p. 420–433, nov. 2012. Disponível em: <<https://academic.oup.com/bfg/article-lookup/doi/10.1093/bfgp/els030>>.

KOHANY, O. et al. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. **BMC bioinformatics**, v. 7, p. 474, jan. 2006. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1634758&tool=pmcentrez&rendertype=abstract>>. Acesso em: 2 jun. 2014.

KOPKA, J. et al. GMD@CSB.DB: the Golm Metabolome Database. **Bioinformatics**, v. 21, n. 8, p. 1635–1638, 15 abr. 2005. Disponível em: <<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti236>>.

KOTSIANTIS, S. B.; ZAHARAKIS, I. D.; PINTELAS, P. E. Machine learning: a review of classification and combining techniques. **Artificial Intelligence Review**, v. 26, n. 3, p. 159–190, 10 nov. 2006. Disponível em: <<http://link.springer.com/10.1007/s10462-007-9052-3>>.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature methods**, v. 9, n. 4, p. 357–9, abr. 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3322381&tool=pmcentrez&rendertype=abstract>>. Acesso em: 10 jul. 2014.

LEE, M. J. et al. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. **Cell**, v. 149, n. 4, p. 780–794, maio 2012. Disponível em: <<http://dx.doi.org/10.1016/j.cell.2012.03.031>>.

LEE, T. I. et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. **Science (New York, N.Y.)**, v. 298, n. 5594, p. 799–804, 25 out. 2002. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1075090>>.

- LEWIS, J. a. et al. Exploiting natural variation in *Saccharomyces cerevisiae* to identify genes for increased ethanol resistance. **Genetics**, v. 186, n. 4, p. 1197–1205, 1 dez. 2010.
- LI, B.; DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC Bioinformatics**, v. 12, n. 1, p. 323, 2011. Disponível em: <<http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-323>>.
- LI, F. et al. The yeast cell-cycle network is robustly designed. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 14, p. 4781–4786, 2004.
- LICATA, L. et al. MINT, the molecular interaction database: 2012 update. **Nucleic Acids Research**, v. 40, n. D1, p. D857–D861, jan. 2012. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr930>>.
- LOHMANN, G. et al. Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. **PLoS ONE**, v. 5, n. 4, 2010.
- LOVE, M. I.; HUBER, W.; ANDERS, S. **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**bioRxiv. [s.l: s.n.]. Disponível em: <<http://biorxiv.org/content/early/2014/05/27/002832.abstract>>.
- LUO, W. et al. GAGE: generally applicable gene set enrichment for pathway analysis. **BMC bioinformatics**, v. 10, n. 1, p. 161, 27 maio 2009. Disponível em: <<http://www.biomedcentral.com/1471-2105/10/161>>.
- LUO, W.; BROUWER, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. **Bioinformatics (Oxford, England)**, v. 29, n. 14, p. 1830–1, 15 jul. 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23740750>>.
- LUSCOMBE, N. M. et al. Genomic analysis of regulatory network dynamics reveals large topological changes. **Nature**, v. 431, n. 7006, p. 308–312, 16 set. 2004.
- MA, M.; LIU, Z. L. **Mechanisms of ethanol tolerance in saccharomyces cerevisiae**Applied Microbiology and Biotechnology, jul. 2010. . Disponível em: <<http://link.springer.com/10.1007/s00253-010-2594-3>>.
- MANGAN, S.; ZASLAVER, A.; ALON, U. The Coherent Feedforward Loop Serves as a Sign-sensitive Delay Element in Transcription Networks. **Journal of Molecular Biology**, v. 334, n. 2, p. 197–204, nov. 2003. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14607112>>.
- MANI, R.-S.; CHINNAIYAN, A. M. Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. **Nature reviews. Genetics**, v. 11, n. 12, p. 819–29, dez. 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21045868>>. Acesso em: 29 out. 2014.
- MARIS, A. F. et al. Diauxic shift-induced stress resistance against hydroperoxides in *Saccharomyces cerevisiae* is not an adaptive stress response and does not depend on

functional mitochondria. **Current Genetics**, v. 39, n. 3, p. 137–149, 16 maio 2001. Disponível em: <<http://link.springer.com/10.1007/s002940100194>>.

MUSSATTO, S. I. et al. **Technological trends, global market, and challenges of bio-ethanol production** *Biotechnology Advances* Elsevier Inc., , nov. 2010. . Disponível em: <<http://dx.doi.org/10.1016/j.biotechadv.2010.07.001>>.

NARAYANAN, A. et al. Defects in Protein Folding Machinery Affect Cell Wall Integrity and Reduce Ethanol Tolerance in *S. cerevisiae*. **Current Microbiology**, v. 73, n. 1, p. 38–45, 18 jul. 2016. Disponível em: <<http://link.springer.com/10.1007/s00284-016-1024-x>>.

NAVARRO-TAPIA, E. et al. Ethanol Cellular Defense Induce Unfolded Protein Response in Yeast. **Frontiers in Microbiology**, v. 7, n. FEB, p. 1–12, 18 fev. 2016. Disponível em: <<http://journal.frontiersin.org/Article/10.3389/fmicb.2016.00189/abstract>>.

NIGAM, P. S.; SINGH, A. Production of liquid biofuels from renewable resources. **Progress in Energy and Combustion Science**, v. 37, n. 1, p. 52–68, fev. 2011. Disponível em: <<http://dx.doi.org/10.1016/j.pecs.2010.01.003>>.

OGATA, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, v. 27, n. 1, p. 29–34, 1 jan. 1999. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/27.1.29>>.

PENG, Y. et al. IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [s.l.: s.n.]6044 LNBIp. 426–440.

PEREIRA, F. B. et al. Genome-wide screening of *Saccharomyces cerevisiae* genes required to foster tolerance towards industrial wheat straw hydrolysates. **Journal of Industrial Microbiology and Biotechnology**, v. 41, n. 12, p. 1753–1761, 2014.

PÉREZ-GALLARDO, R. V. et al. Reactive oxygen species production induced by ethanol in *Saccharomyces cerevisiae* increases because of a dysfunctional mitochondrial iron-sulfur cluster assembly system. **FEMS Yeast Research**, v. 13, n. 8, p. 804–819, dez. 2013. Disponível em: <<https://academic.oup.com/femsyr/article-lookup/doi/10.1111/1567-1364.12090>>.

PIRUZIAN, E. et al. Integrated network analysis of transcriptomic and proteomic data in psoriasis. **BMC systems biology**, v. 4, p. 41, 2010.

PISKUR, J. et al. How did *Saccharomyces* evolve to become a good brewer? **Trends in Genetics**, v. 22, n. 4, p. 183–186, abr. 2006. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0168952506000333>>.

PRONK, J. T.; YDE STEENSMA, H.; VAN DIJKEN, J. P. Pyruvate Metabolism in *Saccharomyces cerevisiae*. **Yeast**, v. 12, n. 16, p. 1607–1633, dez. 1996. Disponível em: <<http://doi.wiley.com/10.1002/%28SICI%291097-0061%28199612%2912%3A16%3C1607%3A%3AAID-YEA70%3E3.0.CO%3B2-4>>.

QUINLAN, A. R.; HALL, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. **Bioinformatics**, v. 26, n. 6, p. 841–842, 2010.

R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna, Austria, 2015. Disponível em: <<http://www.r-project.org>>.

REGULY, T. et al. Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. **Journal of biology**, v. 5, n. 4, p. 11, 2006. Disponível em: <<http://jbiol.com/content/5/4/11>>.

REIMAND, J. et al. g:Profiler—a web server for functional interpretation of gene lists (2016 update). **Nucleic Acids Research**, v. 44, n. W1, p. W83–W89, 8 jul. 2016. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw199>>.

RICHARDS, V. P. et al. Transcriptomic and genomic evidence for *Streptococcus agalactiae* adaptation to the bovine environment. **BMC genomics**, v. 14, p. 920, jan. 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3890567&tool=pmcentrez&rendertype=abstract>>.

RITCHIE, M. D. et al. Methods of integrating data to uncover genotype – phenotype interactions. **Nature Publishing Group**, n. January, 2015. Disponível em: <<http://dx.doi.org/10.1038/nrg3868>>.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139–140, 1 jan. 2010. Disponível em: <<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp616>>.

ROLLAND, T. et al. A Proteome-Scale Map of the Human Interactome Network. **Cell**, v. 159, n. 5, p. 1212–1226, nov. 2014. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0092867414014226>>.

SAID, M. R. et al. Global network analysis of phenotypic effects: protein networks and toxicity modulation in *Saccharomyces cerevisiae*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 52, p. 18006–18011, 28 dez. 2004.

SCHAFFTER, T.; MARBACH, D.; FLOREANO, D. GeneNetWeaver: in silico benchmark generation and performance profiling of network inference methods. **Bioinformatics**, v. 27, n. 16, p. 2263–2270, 15 ago. 2011. Disponível em: <<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr373>>.

SCHAUER, N. et al. GC-MS libraries for the rapid identification of metabolites in complex biological samples. **FEBS Letters**, v. 579, n. 6, p. 1332–1337, 28 fev. 2005. Disponível em: <<http://doi.wiley.com/10.1016/j.febslet.2005.01.029>>.

SONG, G. et al. AGAPE (Automated Genome Analysis PipelinE) for Pan-Genome Analysis of *Saccharomyces cerevisiae*. **PLOS ONE**, v. 10, n. 3, p. e0120671, 17 mar. 2015. Disponível em: <<http://dx.plos.org/10.1371/journal.pone.0120671>>.

STANLEY, D. et al. The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. **Journal of applied microbiology**, v. 109, n. 1, p. 13–24, jul. 2010a. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20070446>>.

STANLEY, D. et al. Transcriptional changes associated with ethanol tolerance in *Saccharomyces cerevisiae*. **Applied Microbiology and Biotechnology**, v. 88, n. 1, p. 231–239, set. 2010b.

STEFANINI, I. et al. A Systems Biology Approach to Dissection of the Effects of Small Bicyclic Peptidomimetics on a Panel of *Saccharomyces cerevisiae* Mutants. **Journal of Biological Chemistry**, v. 285, n. 30, p. 23477–23485, 23 jul. 2010. Disponível em: <<http://www.jbc.org/lookup/doi/10.1074/jbc.M110.125153>>.

SUPEK, F. et al. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. **PLoS ONE**, v. 6, n. 7, p. e21800, 18 jul. 2011. Disponível em: <<https://dx.plos.org/10.1371/journal.pone.0021800>>.

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, n. 5, p. 975–984, 1 maio 2012.

TAKAGI, H.; TAGUCHI, J.; KAINO, T. Proline accumulation protects *Saccharomyces cerevisiae* cells in stationary phase from ethanol stress by reducing reactive oxygen species levels. **Yeast**, v. 33, n. 8, p. 355–363, ago. 2016. Disponível em: <<http://doi.wiley.com/10.1002/yea.3154>>.

TARAILO-GRAOVAC, M.; CHEN, N. Using RepeatMasker to identify repetitive elements in genomic sequences. **Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]**, v. Chapter 4, n. March, p. Unit 4.10, mar. 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19274634>>. Acesso em: 28 maio. 2014.

TEIXEIRA, M. C. et al. Genome-Wide Identification of *Saccharomyces cerevisiae* Genes Required for Maximal Tolerance to Ethanol. **Applied and Environmental Microbiology**, v. 75, n. 18, p. 5761–5772, 15 set. 2009. Disponível em: <<http://aem.asm.org/cgi/doi/10.1128/AEM.00845-09>>.

VALENTE, G. T. et al. The development of a universal in silico predictor of protein-protein interactions. **PloS one**, v. 8, n. 5, p. e65587, jan. 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3669264&tool=pmcentrez&rendertype=abstract>>. Acesso em: 3 nov. 2014.

VERES, D. V. et al. ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. **Nucleic acids research**, v. 43, n. Database issue, p. D485–93, 28 jan. 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25348397>>.

WANG, J. et al. RNA-seq based identification and mutant validation of gene targets related to ethanol resistance in cyanobacterial *Synechocystis* sp. PCC 6803. **Biotechnology for Biofuels**, v. 5, n. 1, p. 89, 2012. Disponível em:

<<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3564720&tool=pmcentrez&rendertype=abstract>>.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews. Genetics**, v. 10, n. 1, p. 57–63, jan. 2009. Disponível em: <<http://www.nature.com/doi/10.1038/nrg2484>>.

WASSERMAN, S.; FAUST, K. **Social Network Analysis: Methods and Applications**. Cambridge: Cambridge University Press, 1994.

WU, T. D.; WATANABE, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. **Bioinformatics**, v. 21, n. 9, p. 1859–1875, 1 maio 2005. Disponível em: <<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti310>>.

XU, K. et al. Path lengths in protein-protein interaction networks and biological complexity. **PROTEOMICS**, v. 11, n. 10, p. 1857–1867, maio 2011. Disponível em: <<http://doi.wiley.com/10.1002/pmic.201000684>>.

XU, X. et al. Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. **BMC bioinformatics**, v. 14 Suppl 9, n. Suppl 9, p. S1, 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3697991&tool=pmcentrez&rendertype=abstract>>.

YANG, S. et al. **Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress** **BMC Genomics** **BMC Genomics**, , 2012. . Disponível em: <[BMC Genomics](#)>.

YANG, S. et al. Systems Biology Analysis of *Zymomonas mobilis* ZM4 Ethanol Stress Responses. **PLoS ONE**, v. 8, n. 7, p. 89, 2013. Disponível em: <<http://www.biotechnologyforbiofuels.com/content/5/1/89>>.

YANG, T.-H. et al. YTRP: a repository for yeast transcriptional regulatory pathways. **Database : the journal of biological databases and curation**, v. 2014, n. November, p. bau014, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24608172>>.

YOSHIKAWA, K. et al. Comprehensive phenotypic analysis for identification of genes affecting growth under ethanol stress in *Saccharomyces cerevisiae*. **FEMS Yeast Research**, v. 9, n. 1, p. 32–44, fev. 2009. Disponível em: <<https://academic.oup.com/femsyr/article-lookup/doi/10.1111/j.1567-1364.2008.00456.x>>.

YU, H. et al. High-Quality Binary Protein Interaction Map of the Yeast Interactome Network. **Science**, v. 322, n. 5898, p. 104–110, 3 out. 2008. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1158684>>.

YU, K. O. et al. Increased ethanol production from glycerol by *Saccharomyces cerevisiae* strains with enhanced stress tolerance from the overexpression of SAGA complex components. **Enzyme and Microbial Technology**, v. 51, n. 4, p. 237–243, set. 2012. Disponível em: <<http://dx.doi.org/10.1016/j.enzmictec.2012.07.003>>.

ZERBINO, D. R.; BIRNEY, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, v. 18, n. 5, p. 821–829, 21 fev. 2008. Disponível em: <<http://www.genome.org/cgi/doi/10.1101/gr.074492.107>>.

ZHANG, Y. et al. The importance of engineering physiological functionality into microbes. **Trends in Biotechnology**, v. 27, n. 12, p. 664–672, dez. 2009.

ZHAO, F. et al. Enhancing *Saccharomyces cerevisiae* reactive oxygen species and ethanol stress tolerance for high-level production of protopanaxadiol. **Bioresource Technology**, v. 227, p. 308–316, mar. 2017. Disponível em: <<http://dx.doi.org/10.1016/j.biortech.2016.12.061>>.

ZYRINA, A. N. et al. Mitochondrial Superoxide Dismutase and Yap1p Act as a Signaling Module Contributing to Ethanol Tolerance of the Yeast *Saccharomyces cerevisiae*. **Applied and Environmental Microbiology**, v. 83, n. 3, p. 1–13, 1 fev. 2017. Disponível em: <<http://aem.asm.org/lookup/doi/10.1128/AEM.02759-16>>.

**APÊNDICE 1 – Tabela completa com Termos do REVIGO**

BMA64-1A Up	BY4742 Up	X2180-1A Up	S288C Up	SEY6210 Up	BY4741 Up
triglyceride metabolism	import into cell	import into cell	response to external stimulus	toxin metabolism	nitrogen compound transport
response to abiotic stimulus	cellular aldehyde metabolism	response to external stimulus	cellular aldehyde metabolism	response to oxidative stress	regulation of biological process
eisosome assembly	response to temperature stimulus	neutral lipid metabolism	import into cell	import into cell	cellular response to stress
regulation of protein stability	protein maturation	protein stabilization	peroxisome organization	homeostatic process	sporulation resulting in formation of a cellular spore
carnitine metabolism	protein stabilization	eisosome assembly	ubiquitin-dependent protein catabolism via the multivesicular body sorting pathway	protein oligomerization	localization
glucose nucleotide 6-phosphate nucleotide metabolism	eisosome assembly	antibiotic catabolism	6-phosphate cellular carbohydrate metabolism	antibiotic catabolism	response to stimulus
chaperone-mediated protein folding	glucose 6-phosphate metabolism	polysaccharide metabolism	cellular carbohydrate metabolism	protein ubiquitination	developmental process
coenzyme metabolism	antibiotic catabolism	catabolism	catabolism	hydrogen sulfide biosynthesis	cellular component organization
catabolism	catabolism	coenzyme metabolism	pyridine-containing compound metabolism	catabolism	biological regulation
polysaccharide metabolism	nucleophagy	nucleophagy	carbohydrate homeostasis	pyridine-containing compound metabolism	cellular protein modification process
protein unfolding	pyridine-containing compound metabolism	protein-cofactor linkage	localization	coenzyme metabolism	growth of unicellular organism as a thread of attached cells
hydrogen peroxide metabolism	response to stimulus	thioester metabolism	response to stimulus	response to stimulus	
protein glutathionylation	signaling	pyridine-containing compound metabolism	signaling	glucose 6-phosphate process metabolism	
carbohydrate derivative metabolism	protein refolding	localization	autophagy	nucleophagy	
reactive oxygen species metabolism	coenzyme metabolism	cell communication	hydrogen peroxide metabolism	protein refolding	
generation of precursor metabolites and energy	localization	receptor metabolism	carbon utilization	autophagy	
energy coupled proton transport, down electrochemical gradient	autophagy	carbohydrate metabolism	cell communication	cell communication	
organic hydroxy compound metabolism	hydrogen peroxide metabolism	response to stimulus	receptor metabolism	hydrogen peroxide metabolism	
cofactor metabolism	thioester metabolism	carbon utilization	generation of precursor metabolites and energy	organic hydroxy compound metabolism	
carbohydrate metabolism	cell communication	sulfur compound metabolism	late nucleophagy	sulfur compound metabolism	
	carbohydrate metabolism	cofactor metabolism	carbohydrate metabolism	carbohydrate metabolism	
	generation of precursor metabolites and energy	autophagy	cofactor metabolism	anatomical structure formation involved in morphogenesis	
	cofactor metabolism	hydrogen peroxide metabolism	coenzyme metabolism	carbohydrate derivative metabolism	
		generation of precursor metabolites and energy		generation of precursor metabolites and energy	
		sporulation resulting in formation of a cellular spore		cofactor metabolism	

BMA64-1A Down	BY4742 Down	X2180-1A Down	S288C Down	SEY6210 Down	BY4741 Down
cell wall macromolecule metabolism	cell wall macromolecule metabolism	cell wall macromolecule metabolism	glycoprotein metabolism	regulation of transferase activity	cytoskeleton organization
chromatin remodeling at centromere	nucleus organization	ATP complex coupled ion transmembrane transport	Golgi organization	localization within membrane	microtubule-based process
protein complex localization	glycosylation	tRNA gene clustering	localization within membrane	cell wall macromolecule metabolism	cellular component organization or biogenesis
hyperosmotic response	retrograde vesicle-mediated transport, Golgi to ER	glycosylation	organophosphate biosynthesis	pyrimidine-containing compound metabolism	biological regulation
cellular alcohol metabolism	anatomical structure homeostasis	positive regulation of GTPase activity	regulation of response to DNA damage stimulus	septin ring organization	response to stimulus
asexual reproduction	proteasomal ubiquitin-independent protein catabolism	establishment or maintenance of cell polarity	establishment or maintenance of cell polarity	glycosylation	response to heat
positive regulation of transferase activity	establishment or maintenance of cell polarity	response to osmotic stress	glycosylation	cell division	small molecule catabolism
establishment of cell polarity	microtubule-based movement	methylation	protein folding	cellular metabolic compound salvage	negative regulation of catalytic activity
microtubule-based movement	protein folding	movement of cell or subcellular component	methylation	chaperone-mediated protein folding	
multicellular organismal process	2-micrometer plasmid partitioning	(1->3)-beta-D-glucan metabolism		methylation	
nitrogen cycle metabolism	movement of cell or subcellular component			(1->3)-beta-D-glucan metabolism	
multicellular organism development	cellular metabolic compound salvage			protein folding	
(1->3)-beta-D-glucan metabolism				carbohydrate derivative metabolism	
				coenzyme metabolism	

**APÊNDICE 2 – Genes Diferencialmente Expressos e Metabólitos Diferencialmente Abundantes**

Estes dados podem ser obtidos mediante solicitação via e-mail à [ivanr.wolf@gmail.com](mailto:ivanr.wolf@gmail.com)

### **APÊNDICE 3 – Estudo Piloto de *Machine Learning* aplicado a redes Biológicas**

## **Estudo piloto: Machine Learning aplicado a redes biológicas**

### **Introdução**

Este apêndice apresenta os resultados obtidos com uma análise de *machine learning*. Estes resultados foram alocados como apêndice uma vez que foi descoberto um problema em uma das bibliotecas de tratamento da linhagem BMA64-1A, a qual poderia adicionar uma forte tendência a subestimar o número de genes diferencialmente expressos comprometendo assim os dados de aprendizado de máquina. Também, as redes utilizadas nesta abordagem eram de dados puramente preditos, e para o estudo final, optou-se por utilizar redes com dados validados experimentalmente (ver **Materiais e Métodos** no texto principal). Além disso, o manuscrito já redigido não contempla os dados de aprendizado de máquina, em virtude de não terem gerado dados tão satisfatórios como esperávamos. Apesar destes inconvenientes os resultados de aprendizado de máquina estão de acordo com o observado no texto principal e por isso foram mantidos nesta tese como um apêndice.

### **Métodos**

#### **Antiga análise do transcriptoma para construção do dataset**

Os *paired-end reads* foram limpos com os seguintes critérios: 1 - remoção de *reads* com adaptadores; 2 – remoção de *reads* com mais de 10% de bases desconhecidas (N); 3 – remoção de *reads* de baixa qualidade (definidas como *reads* com mais de 50% de bases com valor de qualidade  $\leq 10$ ). Após a limpeza as bibliotecas foram mapeadas no genoma de referência da levedura, linhagem S288C, usando o *software* Bowtie2 v2.2.9 (LANGMEAD; SALZBERG, 2012). A abundância dos *reads* foi

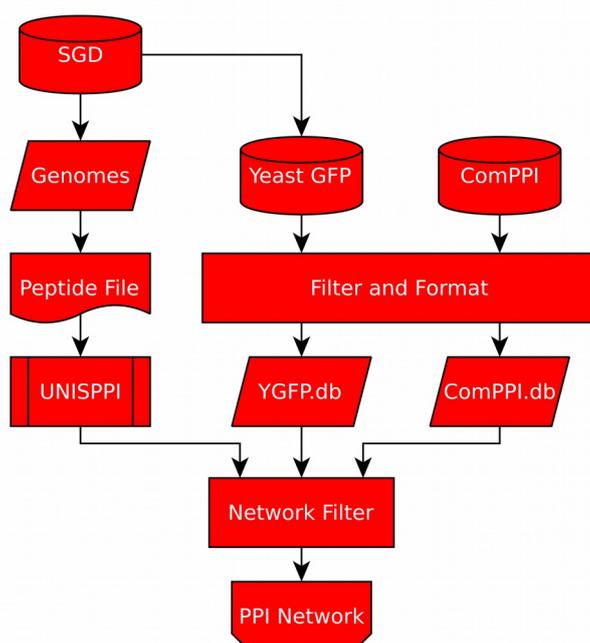
estimada por meio do RSEM v1.3.0 (LI; DEWEY, 2011) e a análise de expressão diferencial foi realizada utilizando o EdgeR v3.12.1 (ROBINSON; MCCARTHY; SMYTH, 2010) do pacote Bioconductor. Genes mostrando diferenças significativas (FDR <0,05) foram considerados diferencialmente expressos.

### Machine learning aplicado a redes biológicas

Para a reconstrução das interações proteína-proteína (PPIs), todas as sequências protéicas das linhagens foram obtidas do *Saccharomyces* Genome Database (SGD) (CHERRY et al., 2012) e todos os pares de proteínas foram submetidos ao preditor UNISPPI (VALENTE et al., 2013), na qual cada par de proteínas foi definido como uma aresta na rede. A porcentagem sobreposta das localizações subcelulares foi obtida a partir do banco de dados ComPPI (VERES et al., 2015) e SGD, a porcentagem de sobreposição de cada base de dados constitui uma nova pontuação para cada par de proteínas. Finalmente, um procedimento de filtro sobre o valor do preditor e a porcentagem de localização sub-celular dos dois bancos de dados usando Análise de Componentes Principais (PCA) foi aplicado para obter uma rede (**Figura 1 desse apêndice**). Adicionalmente, as redes *Yeast Interactome Database* (Y2H-union) (YU et al., 2008); e a rede regulatória utilizada no programa *gene net weather* (GNW) (KIM; IMOTO; MIYANO, 2003; SCHAFFTER; MARBACH; FLOREANO, 2011) foram unidas à rede predita formando uma rede integrada para o ensaio de aprendizado de máquina.

Então, os dados de expressão foram utilizados para construção de duas redes em estados diferentes, controle e tratamento, para cada uma das linhagens. Para isso seguiram-se os procedimentos dos **materiais e métodos** no texto principal do tópico “3.8. Machine Learning aplicado a redes biológicas”. Brevemente, o protocolo realizado segue-

se: 1 - os nós com  $\log_2$  fold-change positivo e significativos eram removidos da rede para gerar a rede controle; 2 - para gerar a rede tratamento todos os nós com  $\log_2$  fold-change negativo e significativos eram removidos e os positivos eram adicionados; 3 - ao final desta etapa uma intersecção das redes das linhagens HT e uma das linhagens LT foi obtida para os estados de tratamento e controle.



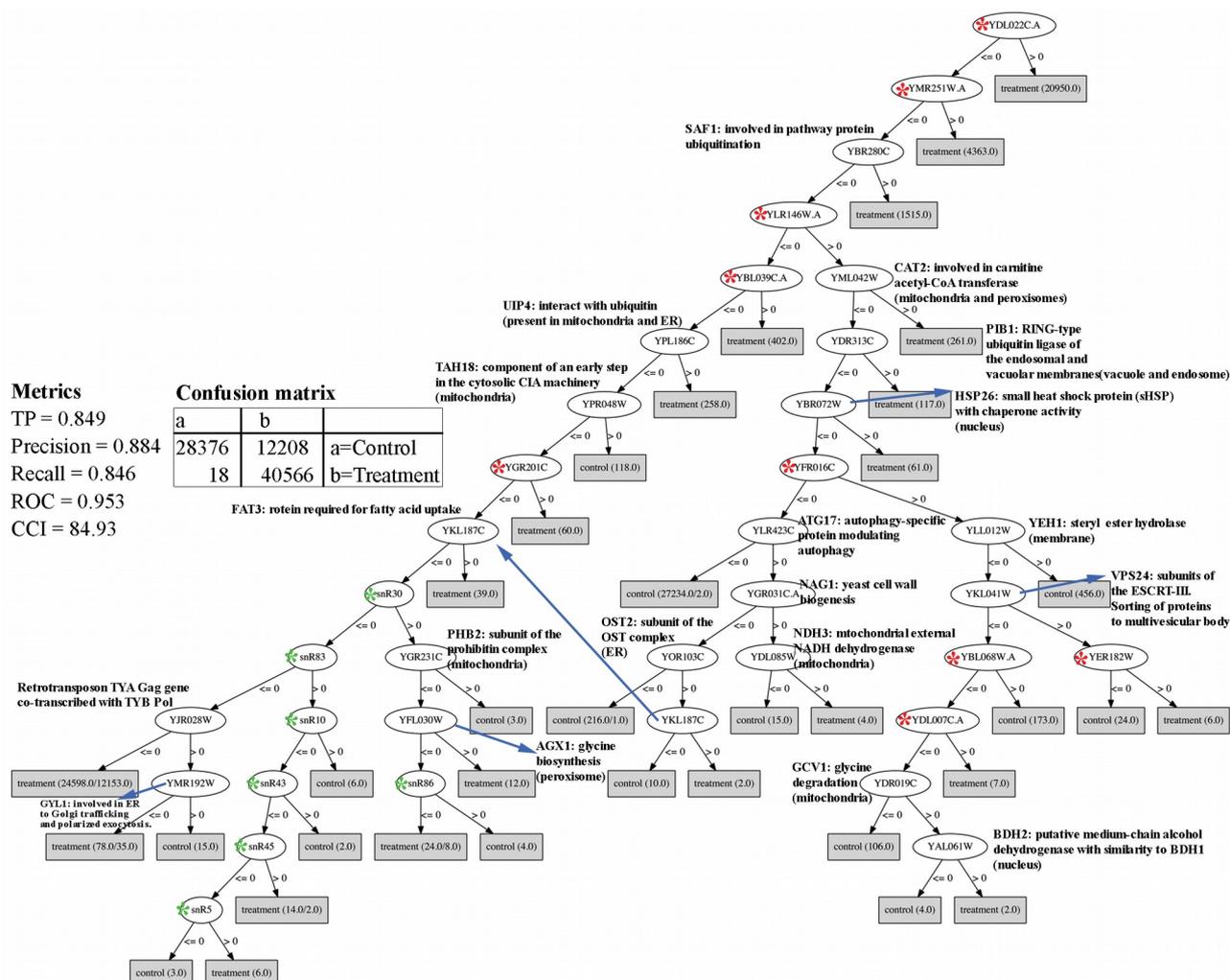
**Figura 1:** Pipeline para construção da rede PPI.

## Resultados

### Machine learning em redes biológicas

O estudo de aprendizado de máquina nas redes (CT, a matriz inteira e HTLT, a matriz de tratamento) reporta desempenhos de treinamento satisfatórios. Para as redes CT, 18 e 3 folhas levam à classe “tratamento” se estiverem conectadas ( $> 0$ ) ou desconectadas ( $\leq 0$ ), respectivamente. Isso significa que durante o tratamento com EtOH, geralmente novas arestas estão presentes na rede. Por outro lado, a árvore de decisão das redes HTLT apresentou um número semelhante de folhas conectadas para

classificar uma instância como HT (9 folhas) ou LT (7 folhas) e 1 folha não conectada para ambas as classes de linhagens (Figura 2 e 3 desse apêndice).



Uma análise mais detalhada da árvore de decisão CT observa-se a presença de muitos genes com funções desconhecidas na rede, sendo então ranqueados como importantes para o fenótipo de tolerância ao EtOH. Para os genes conhecidos, genes envolvidos na estrutura do ribossomo (pequenos RNAs nucleolares), via de ubiquitinação, funcionamento mitocondrial, peroxissomos, membrana e biogênese da parede celular,

proteínas de choque térmico, um gene envolvido na endocitose, autofagia, complexo de Golgi e 1 retrotransposon foram classificados como nós importantes. Alguns genes identificados pela árvore de decisão são ser novos quanto ao estudo de tolerância ao etanol (**Figura 1 desse apêndice**).

### Metrics

TP = 0.850

Precision = 0.885

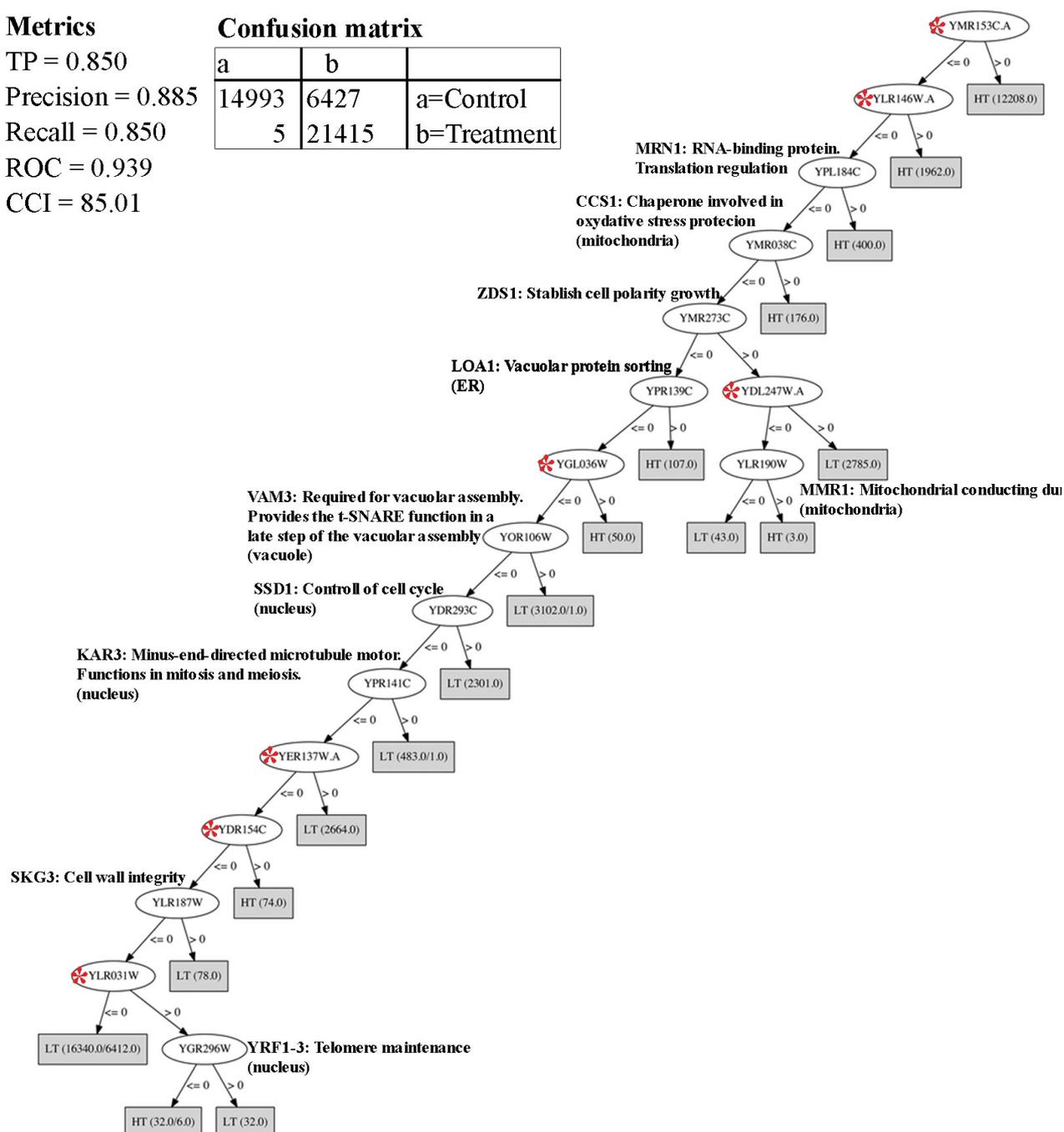
Recall = 0.850

ROC = 0.939

CCI = 85.01

### Confusion matrix

a	b	
14993	6427	a=Control
5	21415	b=Treatment



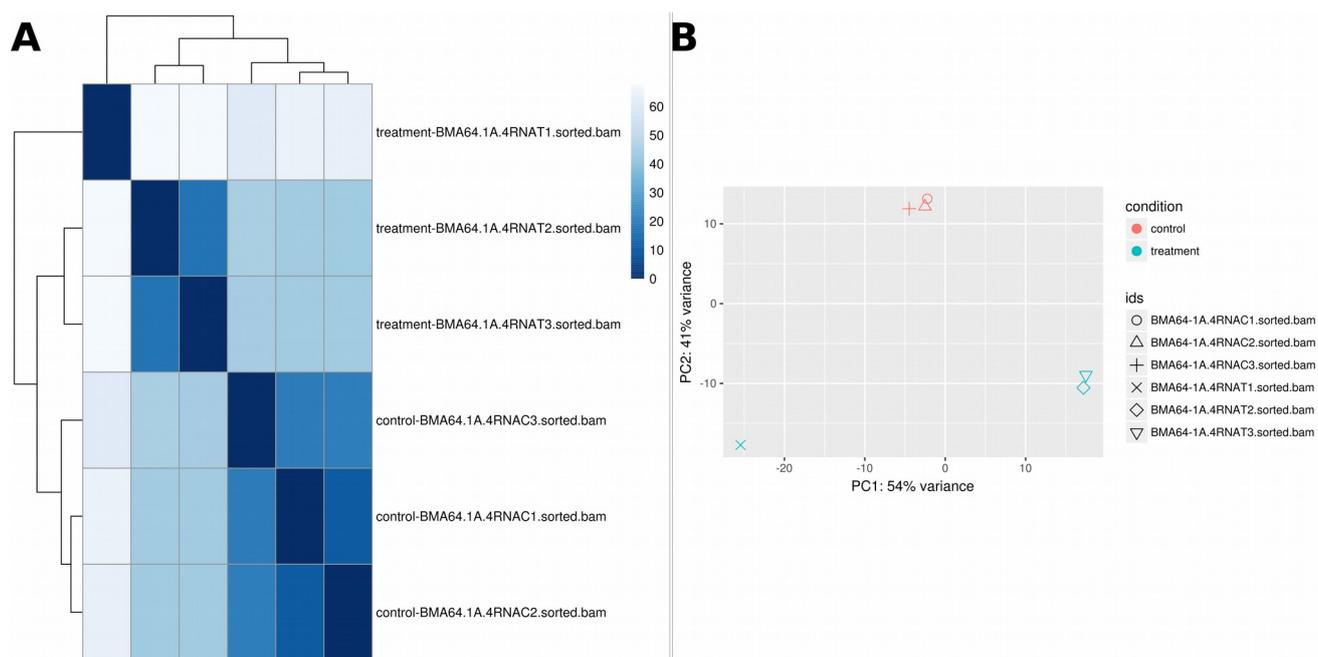
**Figura 3:** Árvore de decisão da matriz HTLT (matriz reportando o tratamento). Nesta árvore é possível observar os nós importantes para classificar os mesmos como relevantes à classe HT ou LT. \* **vermelhos:**

genes com função desconhecida. **TP**: taxa de verdadeiros positivos; **ROC**: *receiver operating characteristic curve*; **CCI**: porcentagem de instâncias corretamente classificadas.

### Experimento de expressão diferencial

Análises mais detalhadas dos resultados do experimento de expressão diferencial possibilitaram observar discrepâncias principalmente com relação ao número de genes diferencialmente expressos. Como exemplo, a linhagem BMA64-1A exibiu um total de 584 genes diferencialmente expressos enquanto que a média para outras linhagens era de 4935. Assim o protocolo do pacote DESeq2 (LOVE; HUBER; ANDERS, 2014) foi utilizado para verificar as causas da discrepância.

A análise de distância entre as bibliotecas e a PCA com os dados de expressão (**Figure 4 desse apêndice**) permitiu identificar que a biblioteca 4RNAT1 não exibiu o comportamento esperado de agrupar com as outras bibliotecas 4RNAT2 e 4RNAT3 (tratamento). O alinhamento da biblioteca também revelou que menos de 50% dos *reads* alinhavam corretamente contra o genoma de BMA64-1A. Assim, foi efetuada a análise do alinhamento da biblioteca para os dados antes e depois da limpeza efetuada pelo Trimmomatic; e também contra possíveis contaminantes (*Escherichia coli*, *Candida albicans* e *Homo sapiens*). Esses resultados revelaram não haver problemas relativos nem a limpeza e nem a presença de contaminantes na amostra.



**Figura 4:** Análise das bibliotecas da linhagem BMA64-1A. **A:** clusterização das bibliotecas utilizando os dados de expressão normalizados; **B:** PCA com os dados de expressão de cada biblioteca.

Após a revisão de todos os métodos de bioinformática aplicados, um relatório foi enviado à empresa responsável pelo sequenciamento, a qual constatou que o problema muito provavelmente é decorrente da degradação de 1 amostra específica e que o re-sequenciamento não seria efetivo para solucionar este problema. Assim, tendo em vista a manutenção das condições experimentais, a resolução do problema foi dada pela remoção desta biblioteca da análise a fim de garantir a qualidade e confiabilidade dos resultados. Este tipo de procedimento é comum e já foi aplicado em trabalhos anteriores que utilizaram o pacote DESeq sem prejuízos ao resultado (ver (RICHARDS et al., 2013)).

## Discussão

Alguns dos genes presentes na árvore de decisão (**Figura 2 desse apêndice**) já foram relatados como genes importantes na tolerância ao etanol, como a HSP26 (expressão detectada apenas em níveis mais elevados de etanol) e VPS24 (proteínas vacuolares essenciais sob estresse ao etanol) (YOSHIKAWA et al., 2009; MA; LIU, 2010;

STANLEY et al., 2010a). Os genes NDE2 e AGX1 também podem estar envolvidos na tolerância ao etanol (KASAVI et al., 2016).

Pelo menos 8 genes identificados na árvore de decisão codificam proteínas mitocondriais e têm um papel importante na tolerância ao etanol. A toxicidade do etanol parece estar relacionada com as ROS nas mitocôndrias (COSTA et al., 1997; PEREIRA et al., 2014). A redução de ROS pode proteger as células de *S. cerevisiae* aumentando sua taxa de sobrevivência (ZHAO et al., 2017). Tomados em conjunto, as organelas relacionadas às ROS podem ser afetadas pelo EtOH, o que pode ser observado pela importância de alguns nós na árvore (**Figura 2 desse apêndice**).

Os snRNAs são os genes mais presentes (7 casos) na árvore de decisão, estes geralmente levam para folhas representando a classe “controle”; Assim, é possível que “perda de ribossomos” seja uma consequência importante do estresse causado pelo EtOH (**Figura 2 desse apêndice**). A análise do transcriptoma reportou que genes relacionados à síntese de ribossomos, ou proteínas, estavam diferencialmente sub-expressos, o que pode ocorrer em situações de estresse por etanol (CHANDLER et al., 2004).

Curiosamente, um retrotransposon está presente na parte inferior da árvore de decisão e a ausência de conexões deste nó leva a classe “tratamento”. Surpreendentemente, é bem documentado que elementos genéticos móveis podem ser ativados em situações de estresse (LEWIS et al., 2010; ASLANKOOHI et al., 2013), o que é diferente do cenário aqui observado. No entanto, esta folha não é pura e podemos ver uma taxa de erro maior nela (veja o valor após “/” na **Figura 2 desse apêndice**).

Em relação à árvore de decisão das redes “HTLT” a maioria dos genes está relacionada ao crescimento celular ou processo de divisão celular (como por exemplo os

genes ZDS1, SSD1, KAR3 e YRF1-3), demonstrando a importância destes processos para diferenciar linhagens LT de HT. A presença de KAR3 (que funciona com microtúbulos durante a divisão celular (PEREIRA et al., 2014)) e YRF1-3 leva a uma folha de LT, o que significa que eles estão conectados nas redes de LT (**Figura 3**). Curiosamente, o YRF1-3 é importante para a manutenção da integridade do DNA (STEFANINI et al., 2010), e este dado está de acordo com os transcriptomas e citometria de fluxo usando o anticorpo H2AX, uma vez que encontramos muitos genes relacionados ao reparo do DNA sub expressos em linhagens LTs além de uma maior quantidade de células acumulando danos de forma significativa. Assim, a presença de YRF1-3 em LTs pode ser um esforço celular para combater a fragmentação do genoma para permitir as divisões celulares, provavelmente trabalhando com KAR3.

## **Conclusões**

Apesar de uma das bibliotecas utilizadas apresentar problemas técnicos, os resultados estão de acordo com o que foi observado posteriormente nos experimentos de expressão diferencial, análises de mapas metabólicos e metabolomas reportados no texto principal. Assim essas evidências constituem informações adicionais que reforçam o que fora discutido nessa tese.

#### **APÊNDICE 4 – Projetos realizados durante o doutorado**

### **Projetos realizados durante o doutorado**

#### **Introdução**

Durante o período do doutorado foi possível desenvolver uma série de projetos em parceria que geraram artigos científicos sem prejuízo para o desenvolvimento do projeto principal desta tese.

#### **Artigos completos publicados em periódicos:**

1. AFONSO CORRÊA, BRUNO; LANZA OKADA, GIOVANA; MANTELLE MELLO MALDONADO, LEONARDO; **RODRIGO WOLF, IVAN**; TARGINO VALENTE, GUILHERME; PLANA SIMÕES, RAFAEL

Geração de redes de co-expressão de genes utilizando informações de expressão diferencial. REVISTA ELETRÔNICA PAULISTA DE MATEMÁTICA. , v.14, p.122 - 131, 2019.

2. LOPES, ERALDO FERREIRA; DA COSTA, JOSEMAR GURGEL; **WOLF, IVAN RODRIGO**; LIMA, JOSÉ PAULO DE ARAÚJO; ASTOLFI-FILHO, SPARTACO

Draft Genome Sequence of *Burkholderia gladioli* Coa14, a Bacterium with Petroleum Bioremediation Potential Isolated from Coari Lake, Amazonas, Brazil. GENOME ANNOUNCEMENTS. , v.6, p.e00301-18 - , 2018.

3. **WOLF, IVAN RODRIGO**; PASCHOAL, ALEXANDRE ROSSI; QUIROGA, CECILIA; DOMINGUES, DOUGLAS SILVA; DE SOUZA, ROGÉRIO FERNANDES; PRETTO-GIORDANO, LUCIENNE GARCIA; VILAS-BOAS, LAURIVAL ANTONIO

Functional annotation and distribution overview of RNA families in 27 *Streptococcus agalactiae* genomes. BMC GENOMICS. , v.19, p.556 - , 2018.

4. CANTÃO, NATÁLIA MIRELE; FOGAÇA DE ALMEIDA, LAUANA; **RODRIGO WOLF, IVAN**; OLIVEIRA ALMEIDA, RODRIGO; ALVES DE ALMEIDA CRUZ, ANDRESSA; NUNES, CAROLINE; BARBOSA, ALEXANDRE NAIME; VALENTE, GUILHERME TARGINO; DE MOURA CAMPOS PARDINI, MARIA INÊS; GROTTTO, REJANE MARIA TOMMASINI

HIV Reverse Transcriptase and Protease Genes Variability Can Be a Biomarker Associated with HIV and Hepatitis B or C Coinfection. *Scientific Reports.* , v.8, p.8280 - , 2018.

5. LOURENCETTI, NATÁLIA MANUELA STROHMAYER; **WOLF, IVAN RODRIGO**; LACERDA, MARIA PRISCILA FRANCO; VALENTE, GUILHERME TARGINO; ZANELLI, CLESLEI FERNANDO; SANTONI, MARIANA MARCHI; MENDES-GIANNINI, MARIA JOSÉ SOARES; ENGUITA, FRANCISCO JAVIER; FUSCO-ALMEIDA, ANA MARISA  
Transcriptional profile of a bioethanol production contaminant *Candida tropicalis*. *AMB Express.* , v.8, p.166 - , 2018.

6. GOUVEIA, JUCELI GONZALEZ; **WOLF, IVAN RODRIGO**; VILAS-BOAS, LAURIVAL ANTONIO; HESLOP-HARRISON, JOHN SEYMOUR; SCHWARZACHER, TRUDE; DIAS, ANA LÚCIA  
Repetitive DNA in the Catfish Genome: rDNA, Microsatellites, and Tc1-Mariner Transposon Sequences in *Imparfinis* Species (Siluriformes, Heptapteridae). *Journal of Heredity.* , v.6, p.650 - 657, 2017.

7. GOUVEIA, JUCELI GONZALEZ; **WOLF, IVAN RODRIGO**; DE MORAES-MANÉCOLO, VIVIAN PATRÍCIA OLIVEIRA; BARDELLA, VANESSA BELLINE; FERRACIN, LARA MUNIQUE; GIULIANO-CAETANO, LUCIA; DA ROSA, RENATA; DIAS, ANA LÚCIA  
Isolation and characterization of 5S rDNA sequences in catfishes genome (Heptapteridae and Pseudopimelodidae): perspectives for rDNA studies in fish by C0t method. *Cytotechnology.* , v.June, p.1 - 10, 2016.

#### **Artigos submetidos para periódicos:**

1. MARIANA GATTO, M.D. ; TAÍS FUKUTA DA CRUZ; **IVAN RODRIGO WOLF**; PATRÍCIA APARECIDA BORIM; ALINE MARCIA MARQUES BRAZ; BÁRBARA CASELLA AMORIM; GUILHERME TARGINO VALENTE; MARJORIE ASSIS GOLIM; JAMES VENTURINI; JOÃO PESSOA ARAUJO JUNIOR; ALESSANDRA PONTILLO; ALEXANDRINA SARTORI.

Transcriptional analysis of THP-1 cells infected with *Leishmania infantum* indicates no activation of the inflammasome platform. *PLOS Neglected Tropical Diseases*.

2. FLAVIA HEBELER BARBOSA, PH.D; **IVAN RODRIGO WOLF**; GUILHERME TARGINO VALENTE; ADRIANA CAMARGO FERRASI; FRANCISCO CAMPELLO DO AMARAL MELLO; ELISABETH LAMPE; MARIA INÊS DE M.C. PARDINI; REJANE MARIA TOMMASINI GROTTTO.

New method for next generation sequencing of the hepatitis B virus full genome from clinical specimen: impact for virus genotyping. *PLOS ONE*.

3. LEONARDO N MORAES; REJANE MARIA T GROTTTO; GUILHERME T VALENTE; HELOISA C SAMPAIO; ANGELO J MAGRO; LAUANA FOGAÇA; **IVAN R WOLF**; DAVID PERAHIA; GIOVANNI F SILVA RAFAEL PLANA SIMÕES.

A novel molecular mechanism to explain mutations of the HCV protease associated with resistance against covalently bound inhibitors. *BBA - General Subjects*.

4. MARIANA L. C. ARCURI; LARISSA C. FIALHO; ALESSANDRA VASCONCELLOS NUNES LAITZ; MARIA CECÍLIA P. FUCHS-FERRAZ; **IVAN RODRIGO WOLF**; GUILHERME TARGINO VALENTE; CELSO L. MARINO; IVAN G. MAIA.

Genome-wide identification of multifunctional laccase gene family in *Eucalyptus grandis*: potential targets for lignin engineering and stress tolerance. *Functional & Integrative Genomics*.

5. RAFAEL PLANA SIMÕES, **IVAN R. WOLF**, GUILHERME T. VALENTE

Uncovering patterns of the evolution of genomic entropy and complexity. *Heredity*.

**Artigos redigidos em revisão textual ou em preparação:**

1. **IVAN R. WOLF**; GUILHERME T. VALENTE

Gene Regulatory Networks: Functional and Topological overview of Regulators and Targets.

2. ERIC KAZUO KAWAGOE, GUILHERME TARGINO VALENTE, **IVAN RODRIGO WOLF**

Differential expression analysis for biomarker discovery in ovarian carcinomas.

3. CAMILA C. ALVES, BRUNO F. GAMBA, **IVAN R. WOLF**, LUCILENE RIBEIRO-BICUDO, GUILHERME T. VALENTE.

Machine Learning-Based Methodology for the Clinical Stratification of 22Q11 Deletion Syndrome.

4. CAMILA C. ALVES, **IVAN R. WOLF**, BRUNO F. GAMBA, LUCILENE RIBEIRO-BICUDO, GUILHERME T. VALENTE

A System Biology Approach In The Study Of The 22q11 Deletion Syndrome.

5. **IVAN R. WOLF**, RAFAEL PLANA SIMÕES, GUILHERME T. VALENTE

Co-expression Networks: Biological meaning of Edges.

6. LUIZ HENRIQUE CARDOSO, **IVAN R. WOLF**; GUILHERME T. VALENTE

Co-expression network machine learning analysis of fitness in *Saccharomyces cerevisiae*

7. GIRLÂNIO HOLANDA DA SILVA, **IVAN R. WOLF**, GUILHERME TARGINO VALENTE, EDVALDO APARECIDO AMARAL DA SILVA

Re-Establishment Of Desiccation Tolerance In Germinated Seeds Of *Solanum lycocarpum*

A. St.-Hil.

8. Estudo sobre identificação de RNAs longos não codificantes em tomates, em parceria com o professor Fabio Tebaldi Silveira Nogueira da ESALQ de Piracicaba.

9. Estudo sobre proteínas relacionadas à cicatrização, em parceria com a pesquisadora Lucilene Delazari dos Santos do CEVAP da UNESP Botucatu.

10. Estudo sobre os efeitos de cromossomos B sobre RNAs não codificantes longos no peixe *Astatotilapia latifasciata*, em parceria com a aluna de doutorado Érica Ramos sob orientação do professor Cesar Martins da UNESP Botucatu.

11. Análises em larga escala de proteínas e construção de redes biológicas com foco em estudos de cromossomos B, em parceria com o aluno de doutorado Rafael Takahiro Nakajima sob orientação do professor César Martins da UNESP Botucatu.

12. Análises em larga escala aplicadas ao cromossomo B de *Astyanax sp.*, em parceria com o aluno de doutorado Syed Farhan Ahmad sob orientação do professor César Martins da UNESP Botucatu.