

OSMAR VAZ DE CARVALHO NETTO

**“FISIOLOGIA DA FERMENTAÇÃO ALCOÓLICA: ANÁLISE
DA EXPRESSÃO GÊNICA DE UMA LINHAGEM INDUSTRIAL
DE *Saccharomyces cerevisiae* DURANTE O PROCESSO
FERMENTATIVO”**

**CAMPINAS
2012**

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA



OSMAR VAZ DE CARVALHO NETTO

**“FISIOLOGIA DA FERMENTAÇÃO ALCÓOLICA:
ANÁLISE DA EXPRESSÃO GÊNICA DE UMA LINHAGEM
INDUSTRIAL DE *Saccharomyces cerevisiae* DURANTE O
PROCESSO FERMENTATIVO”**

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
OSMAR VAZ DE CARVALHO NETTO
e aprovada pela Comissão Julgadora.

Tese apresentada ao Instituto de
Biologia para obtenção do Título
de Doutor em Genética e Biologia
Molecular, na área de Genética de
Microrganismos.

Orientador: Prof. Dr. Gonçalo Amarante Guimarães Pereira.

Co-Orientador: Prof. Dr. Juan Lucas Argueso

CAMPINAS

2012

FICHA CATALOGRÁFICA ELABORADA POR
ROBERTA CRISTINA DAL' EVEDOVE TARTAROTTI – CRB8/7430
BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

C253f Carvalho Netto, Osmar Vaz de, 1981-
 Fisiologia da fermentação alcoólica: análise da
 expressão gênica de uma linhagem industrial de
 Saccharomyces cerevisiae durante o processo
 fermentativo / Osmar Vaz de Carvalho Netto. –
 Campinas, SP: [s.n.], 2012.

 Orientador: Gonçalo Amarante Guimarães Pereira.
 Coorientador: Juan Lucas Argueso.
 Tese (doutorado) – Universidade Estadual de
 Campinas, Instituto de Biologia.

 1. *Saccharomyces cerevisiae*. 2. Bioetanol. 3.
 Fermentação alcoólica. 4. Expressão gênica. 5.
 Engenharia genética. I. Pereira, Gonçalo Amarante
 Guimarães. II. Argueso, Juan Lucas. III. Universidade
 Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

Título em Inglês: Physiology of the alcoholic fermentation: gene expression analyses of a *Saccharomyces cerevisiae* industrial strain during the fermentative process

Palavras-chave em Inglês:

Saccharomyces cerevisiae

Bioethanol

Alcoholic fermentation

Gene expression

Genetic engineering

Área de concentração: Genética de Microorganismos

Titulação: Doutor em Genética e Biologia Molecular

Banca examinadora:

Gonçalo Amarante Guimarães Pereira [Orientador]

Pedro de Oliva Neto

Iran Malavazi

Marcelo Menossi Teixeira

Jorge Mauricio Costa Mondego

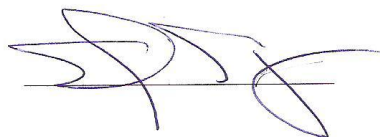
Data da defesa: 31-07-2012

Programa de Pós Graduação: Genética e Biologia Molecular

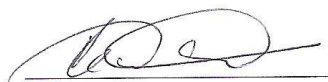
Campinas, 31 de julho de 2012

BANCA EXAMINADORA

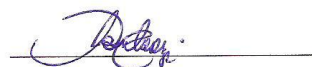
Prof. Dr. Gonçalo Amarante Guimarães Pereira (Orientador)



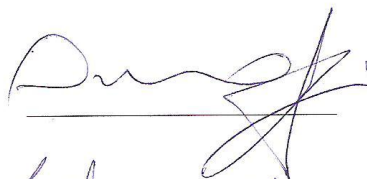
Prof. Dr. Pedro de Oliva Neto



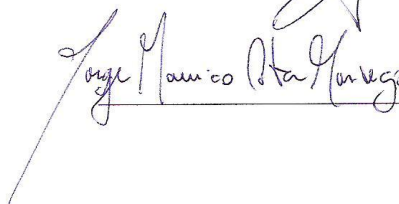
Prof. Dr. Iran Malavazi



Prof. Dr. Marcelo Menossi Teixeira



Prof. Dr. Jorge Mauricio Costa Mondego



Prof. Dr. Adriana Franco Paes Leme



Prof. Dr. Marco Aurélio Takita



Prof. Dr. Marcelo Brocchi



**Dedico esta tese a minha noiva e companheira, Dri, aos meus queridos irmãos
Fernando e Mariana e a todos meus familiares.**

AGRADECIMENTOS

Primeiramente, gostaria de um agradecimento especial ao meu orientador, prof. Dr. Gonçalo Pereira, por todas as oportunidades e pela confiança depositada.

Gostaria de agradecer também de maneira particular ao meu co-orientador, prof. Dr. Lucas Argueso, pela oportunidade em desenvolver parte deste trabalho em seu laboratório nos EUA e pela grande atenção dada ao longo de todos estes anos no desenvolvimento desta tese.

A minha noiva, Dri, presente em todos os momentos com sua energia única e carinho que sempre me motivaram.

A todos meus familiares e amigos, que sempre me motivaram.

Ao prof. Dr. Luiz Basso e sua equipe, ao Dr. Sílvio Andrietta, ao prof. Dr. Piotr Mieczkowski e ao prof. Dr. André Alcarde, pelo auxílio na geração dos dados aqui apresentados.

A toda a equipe do Laboratório de Genômica e Expressão (LGE), em especial a equipe do grupo da Levedura (Sílvia, Felipe, Marcelo, Fabiana, Aline, Ane, Luciana, Melline e Nádia) e a todos aqueles que de alguma forma participaram na obtenção desses dados, em especial aos amigos Eduardo, Luige, Paulo e Gleidson.

A toda a equipe de apoio do LGE, Eliane, Welbe e Ernestina, pelo apoio nos mais diversos assuntos ao longo dos meus estudos.

A todos os membros de minha qualificação, pré-banca e banca, pelas sugestões e contribuições neste trabalho.

Ao Instituto de Biologia da Unicamp, Colorado State University e University of North Carolina at Chapel Hill, pela estrutura disponibilizada.

Às agências e empresas que financiaram este trabalho e minha bolsa de estudos, em especial FAPESP (processo 2008/51500-5) e ETH Bioenergia S/A.

SUMÁRIO

ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABELAS.....	IX
RESUMO	X
ABSTRACT	XI
INTRODUÇÃO	1
1. A LEVEDURA <i>SACCHAROMYCES CEREVISIAE</i>	1
2. TENDÊNCIAS DO MERCADO MUNDIAL DE BIOETANOL	2
3. A PRODUÇÃO DE BIOETANOL NO BRASIL	3
4. MELHORAMENTO GENÉTICO DE LEVEDURAS.....	5
5. MAQUINARIA E TÉCNICAS MOLECULARES PARA ESTUDO DE EXPRESSÃO GÊNICA DE <i>S.</i> <i>CEREVISIAE</i>	5
6. GENÉTICA DE LINHAGENS DE <i>S. CEREVISIAE</i> BRASILEIRAS	8
OBJETIVOS	10
APRESENTAÇÃO DO TRABALHO.....	11
CAPÍTULO 1	13
CAPÍTULO 2	47
CAPÍTULO 3	68
CAPÍTULO 5	141
CONCLUSÕES GERAIS	160
REFERÊNCIAS BIBLIOGRÁFICAS.....	162

ÍNDICE DE FIGURAS

CAPÍTULO 1

Figure 1. Comparative phenotypic analysis of JAY270 and S288C	38
Figure 2. Genetic analysis of cell mass accumulation and ethanol production kinetics	39
Figure 3. Molecular karyotype and gene copy number variation in JAY270	40
Figure 4. Molecular karyotype of meiotic products and segregation of chromosomal rearrangements	41
Figure 5. Band-array analysis of Chr6L and Chr6S homologs in JAY270	42
Figure 6. Genome rearrangements near the ends of chromosomes	43
Figure 7. Phylogenetic placement and high-temperature growth (HTG) alleles in JAY291	44
Figure 8. Structural diversity in <i>S. cerevisiae</i> : rigid and plastic domains of the genome	45

CAPÍTULO 2

Figure 1. Length polymorphisms within coding sequences	62
Figure 2. PCR amplification of length-polymorphic loci	63
Figure 3. Analysis of heterozygous length polymorphisms in JAY270/PE-2	64

CAPÍTULO 3

Figura 1. Processo de domesticação da linhagem JAY270	108
Figura 2. Construção dos cassetes de integração (Hph e Core) utilizados durante o processo de domesticação da linhagem JAY270	109
Figura 3. Estratégia referente à construção do fragmento homólogo de reparo, bem como os produtos de PCR, de digestão e de ligação resultantes	110
Figura 4. Localização dos <i>primers</i> no genoma da levedura <i>Saccharomyces cerevisiae</i> utilizados para confirmar as transformações durante as etapas de domesticação da JAY270	111
Figura 5. Produtos de PCR referentes à confirmação do término da etapa 01 do processo de domesticação da JAY270	111
Figura 6. Produtos de PCR referentes à confirmação do término da etapa 02 do processo de domesticação da JAY270	112
Figura 7. Produtos de PCR referentes à confirmação do término da etapa 03 do processo de domesticação da JAY270	112
Figura 8. Produtos de PCR dos quatro esporos de uma tétrade da linhagem FGY030, corroborando o término da etapa 03 do processo de domesticação da JAY270	113
Figura 9. Produtos de PCR referentes à confirmação do término da etapa 04 do processo de domesticação da JAY270	113
Figura 10. Esquema da estratégia para a confirmação da etapa 05 do processo de domesticação da JAY270	114

Figura 11. Produtos de PCR referentes à confirmação da etapa 05 do processo de domesticação da JAY270	114
Figura 12. Produtos de digestão da PCR referentes à confirmação da etapa 05 do processo de domesticação da JAY270	115
Figura 13. Produtos de PCR para corroborar a permanência de heterozigosidade na linhagem FGY050 e, assim, o término da etapa 05 do processo de domesticação da JAY270.....	115
Figura 14. Gel de agarose da digestão do fragmento de 914pb obtido previamente	116
Figura 15. Representação virtual de parte do cromossomo V da levedura <i>Saccharomyces cerevisiae</i> , segundo o banco de dados SGD	116
Figura 16. Representação virtual de parte do cromossomo V da levedura <i>Saccharomyces cerevisiae</i> (SGD)	117
Figura 17. Representação virtual de parte do cromossomo V da levedura <i>Saccharomyces cerevisiae</i> (SGD)	117

CAPÍTULO 4

Figure 1. Kinetic of the major compounds identified during fermentations	136
Figure 2. Profile of the differentially expressed (DE) genes and a correlation with the content of organic acids.....	137
Figure 3. Gene expression comparison between typical and flocculated fermentations	138
Figure 4. Frequency of allelic expression and pattern of enzymatic digestion.....	139

CAPÍTULO 5

Figura 1. Passos utilizados para a construção do plasmídeo OVP4.....	158
Figura 2. Alteração da região promotora de um gene de interesse	158
Figura 3. Razão do número de células identificadas nos ensaios de fermentação sob competição	159

ÍNDICE DE TABELAS

CAPÍTULO 1

Table 1. Summary of JAY291 genome sequencing and assembly parameter	46
--	----

CAPÍTULO 2

Table 1. List of genes, PCR products, and primers flanking the length polymorphic regions.....	65
Table 2. Analysis of pcr length polymorphic alleles in common sugarcane yeast strains	66
Table 3. PCR monitoring of yeast populations in six bioethanol distilleries during the 2011 season	67

CAPÍTULO 3

Tabela 1. <i>Primers</i> (JAO286; JAO287 e JAO288) para a construção dos cassetes de integração Hph e Core	85
Tabela 2. <i>Primers</i> para confirmar as transformações	87
Tabela 3. Sequência dos <i>primers</i> JAO311 e JAO312 desenhados a confirmação da etapa 05 da domesticação da JAY270	97

CAPÍTULO 4

Table 1. Number of differentially expressed (DE) genes during industrial bioethanol fermentation in two distinct conditions and a comparative analysis among them	140
--	-----

CAPÍTULO 5

Tabela 1. Valores de expressão gênica e descrição resumida da função de dez genes selecionados para estudos de manipulação genética	154
Tabela 2. Promotores utilizados nos ensaios de manipulação genética da linhagem PE-2	155
Tabela 3. Combinação dos promotores utilizados nos ensaios de manipulação gênica para os dez genes selecionados	155
Tabela 4. Linhagens derivadas de PE-2 desenvolvidas nos ensaios de manipulação genética utilizando 10 genes selecionados	156
Tabela 5. Rendimento fermentativo das linhagens geneticamente modificadas em comparação com o isolado selvagem	157

RESUMO

A fermentação alcoólica realizada por leveduras da espécie *Saccharomyces cerevisiae* é atualmente o principal processo em escala econômica para a produção de etanol a partir de fontes renováveis (bioetanol). Entretanto, as linhagens isoladas há quase duas décadas vêm perdendo gradualmente seu desempenho fermentativo ao longo dos anos, reduzindo o rendimento e a estabilidade do processo. Um dos fatores que contribuem para tais perdas está associado às significativas alterações operacionais do processo nestas duas décadas. Aliado a isto, pouco se conhece sobre as bases genéticas das principais linhagens comerciais utilizadas no Brasil, inviabilizando o sucesso de programas de melhoramento genético. Neste ponto, este trabalho propõe de forma inédita um estudo completo de caracterização, análise de expressão e manipulação gênica de PE-2, uma das principais linhagens usadas no processo Brasileiro. Para isto, dados da estrutura, composição e regulação do genoma foram gerados. A partir dessas informações foi desenvolvida, através de manipulações genéticas, uma linhagem auxotrófica para uracila (plataforma biológica) e outras com características de fermentação superiores ao isolado selvagem. O conceito de que o sucesso de manipulações pontuais depende de um conhecimento prévio detalhado da linhagem de interesse, compreendendo tanto suas bases genéticas quanto fisiológicas, foi validado neste estudo. Os dados aqui apresentados abrem novas perspectivas para o desenvolvimento de uma segunda geração de linhagens de *S. cerevisiae* ainda mais robustas, voltadas não apenas para a produção de bioetanol, como também para utilização em outros processos biotecnológicos.

ABSTRACT

The alcoholic fermentation performed by yeasts of the specie *Saccharomyces cerevisiae* is the main current process to produce ethanol from renewable resources (bioethanol) under economic scale. However, strains isolated two decades ago have been gradually losing their original fermentative fitness during the years, decreasing the process yield and stability. One of the factors that contribute for these losses is associated with the meaningful process operational changes along these years. In addition to that, there is a knowledge gap regarding the genetic bases of the commercial strains used on Brazil, not leading to success of the genetic improvements programs. Hereupon, this work propose in a first time way a complete characterization, expression analysis and genetic manipulation study of PE-2, one of the most important strain used on Brazilian process. To this, structure, composition and regulation data of the genome were produced. From such information, it was developed by genetic manipulations assays an auxotrophic yeast for uracile (biological plataform) and others strains with superior fermentative characteristics in comparison with the wild isolated. The concept that the success of precise manipulations depends on the previous detailed knowledge of the interest strain, regarding the genetic and physiology bases, was confirmed herein. The data presented here also establish new perspectives for the construction of a second generation strains even more robusts, designed not only for the bioethanol production, but also to be applied on different biotechnological processes.

DECLARAÇÃO

Declaro para os devidos fins que o conteúdo de minha dissertação de Mestrado/tese de Doutorado intitulada FISILOGIA DA FERMENTAÇÃO ALCÓOLICA: ANÁLISE DA EXPRESSÃO GÊNICA DE UMA LINHAGEM INDUSTRIAL DE *Saccharomyces cerevisiae* DURANTE O PROCESSO FERMENTATIVO:

() não se enquadra no § 3º do Artigo 1º da Informação CCPG 01/08, referente a bioética e biossegurança.

Tem autorização da(s) seguinte(s) Comissão(ões):

(X) CIBio – Comissão Interna de Biossegurança , projeto No. 2011/03, Instituição: UNICAMP – Instituto de Biologia (IB).

() CEUA – Comissão de Ética no Uso de Animais , projeto No. _____, Instituição: _____

() CEP - Comissão de Ética em Pesquisa, protocolo No. _____, Instituição: _____

** Caso a Comissão seja externa ao IB/UNICAMP, anexar o comprovante de autorização dada ao trabalho. Se a autorização não tiver sido dada diretamente ao trabalho de tese ou dissertação, deverá ser anexado também um comprovante do vínculo do trabalho do aluno com o que constar no documento de autorização apresentado.*


Aluno: Osmar Vaz de Carvalho Netto


Orientador: Gonçalo Amante Guimarães Pereira

Para uso da Comissão ou Comitê pertinente:

☒ Deferido () Indeferido

Carimbo e assinatura


Prof. Dr. MARCELO LANCELLOTTI
Presidente da Comissão Interna de Biossegurança
Instituto de Biologia - UNICAMP

Para uso da Comissão ou Comitê pertinente:

() Deferido () Indeferido

Carimbo e assinatura

INTRODUÇÃO

1. A levedura *Saccharomyces cerevisiae*

Saccharomyces cerevisiae é um organismo eucarionte e unicelular. Este fungo é classificado como levedura, assim como uma diversidade de outras 1500 espécies distribuídas em 100 diferentes gêneros. A linhagem padrão de laboratório S288c foi o primeiro eucarioto a ter o genoma sequenciado, revelando uma estrutura composta por aproximadamente 6000 genes distribuídos em 16 cromossomos (Goffeau et al., 1996). A quantidade destes cromossomos nas células pode variar sob a forma haplóide (16 cromossomos) ou diplóide, quando a célula possui dois pares de cada um dos cromossomos (32 no total).

A reprodução destas leveduras ocorre por um evento de duplicação de cromossomos seguido de divisão celular (mitose), em um processo conhecido como brotamento que mantém o número original de cromossomos da célula mãe. Além disso, uma levedura diplóide pode também ser originada através do cruzamento entre dois isolados haplóides. Estes haplóides devem obrigatoriamente ser opostos quanto à diferenciação sexual, isto é, uma célula deve possuir o sexo “a” e outra o “alfa”, gerando uma célula a/alpha. Por outro lado, um haplóide pode ser originado por uma célula diplóide através de um evento de meiose, que produz quatro gametas haplóides (esporos), sendo dois “a” e outros dois “alfa”. Culturas puras de cada um destes esporos podem se manter haplóides quando a linhagem possuir a característica heterotática, ou seja, não sendo capaz de mudar o sexo. Entretanto, isto não ocorre em linhagens homotáticas, onde a cultura adquire a configuração diplóide através do cruzamento de seus esporos irmãos. Linhagens heterotáticas são particularmente interessantes do ponto de vista de manipulação genética, devido à facilidade de se trabalhar com organismos haplóides.

Devido ao reduzido tamanho de seu genoma (12 milhões de pares de bases), diversidade de técnicas disponíveis e facilidade de manipulação, este microrganismo se tornou um importante modelo para estudos de biologia molecular. Além destas aplicações, a capacidade de realizar a bioconversão de açúcares em etanol e dióxido de carbono (fermentação alcoólica) é uma das mais relevantes características dessa espécie. Além disso, estas leveduras são altamente versáteis, sendo utilizadas em uma variedade de processos biotecnológicos que envolvam fermentação. O desenvolvimento de novas técnicas, tanto no processo de fermentação como na biologia molecular, permitiram que as células de levedura se tornassem verdadeiros biorreatores para a produção de uma diversidade de produtos. Entre eles, pode-se exemplificar a produção de proteínas heterólogas (Porro et al., 2011), ácidos

orgânicos (Abbott et al., 2009b), glicerol (Cordier et al., 2007) e mais proeminentemente a produção de biomassa e etanol. Este último vem ganhando uma grande notoriedade mundial, não apenas pela importância na indústria de bebidas (vinho, sake, cerveja, cachaça, entre outros), mas principalmente como fonte de biocombustível (bioetanol).

2. Tendências do mercado mundial de bioetanol

O mercado de leveduras destinadas à produção de biocombustíveis é impulsionado pela crescente demanda mundial por combustíveis renováveis, em especial pelo bioetanol. Este biocombustível pode ser produzido a partir da fermentação de diferentes matérias-primas: milho, cevada, trigo, beterraba, cana-de-açúcar e, mais recentemente, material celulósico hidrolisado (Basso et al., 2010). Em 2010, Estados Unidos e Brasil foram responsáveis por 57% e 33%, respectivamente, dos 86 bilhões de litros de bioetanol produzidos mundialmente (Shrank and Farahmand, 2011).

No Brasil, o aumento contínuo da frota de carros movido a bicomcombustíveis e a demanda por exportação (para ser misturado à gasolina utilizada nos EUA) favorecerá a expansão da cadeia produtiva até 2025, que conta hoje com aproximadamente 450 usinas instaladas. Projeções mais ambiciosas (substituir 10% da gasolina mundial) exigirão um aumento na produção de 10 vezes neste mesmo período (Leite, 2009). Para alcançar essas metas, o país tem investido na melhoria do processo convencional (i.e., etanol de primeira geração) e ao mesmo tempo no desenvolvimento de novas tecnologias para produzir biocombustíveis a partir de biomassa celulósica (etanol celulósico – segunda geração).

Nos EUA, o volume de combustíveis renováveis deverá atingir 136 bilhões de litros até 2022, o que corresponde a 30% de todo o volume de combustíveis do país (Bracmort, 2012). Deste, o Governo Americano propõe que 45% deverão ser provenientes de etanol celulósico. Para isso, em 10 anos o volume total de etanol celulósico deverá ser elevado mil vezes para suprir as exigências do Governo. A União Europeia também estabelece metas de utilização de combustíveis provenientes de fontes renováveis. Em 2020, a expectativa é de que a frota de automóveis deva ser abastecida com pelo menos 10% de biocombustíveis.

Neste ponto, as leveduras são componentes essenciais para ambos os sistemas de produção de bioetanol. Por esta razão, o aumento do número de usinas (para primeira geração) e a implementação de novas tecnologias (para a segunda geração) aumentarão consequentemente a demanda por leveduras já utilizadas e a necessidade de desenvolvimento de linhagens adequadas aos novos processos.

Assim como o setor energético, o setor químico também investe em alternativas sustentáveis, na chamada “Química Verde”. Esta consiste no desenvolvimento de processos e produtos a partir de matéria-prima renovável ou resíduos altamente disponíveis. Neste contexto, existe um mercado emergente de leveduras, com uma demanda em potencial por leveduras capazes de converter matéria-prima renovável em diferentes “compostos químicos verdes”.

3. A produção de bioetanol no Brasil

O processo de produção de bioetanol no Brasil é o mais economicamente viável. Por um lado, possui enorme capacidade produtiva, devido aos baixos custos de obtenção da matéria-prima, às condições geográficas e climáticas (Wheals et al., 1999). Por outro lado, com grande contribuição do Proálcool (Programa Nacional do Álcool) iniciado na década de 70, desenvolveu toda uma tecnologia de produção, distribuição e utilização de bioetanol (Goldemberg, 2008). Com isso, o Brasil é o mais eficiente produtor mundial e, em vista da necessidade global por combustíveis alternativos, se apresenta como um dos países capazes de liderar tal mudança de matriz no mundo.

A obtenção do bioetanol se dá pela via fermentativa em dornas nas quais são adicionados a mistura de mosto (caldo ou melaço de cana-de-açúcar) e uma elevada concentração de leveduras (10 – 15%, m/v) (Amorim et al., 2011). Após o término da fermentação (8 a 15 horas), todo o conteúdo é centrifugado. O vinho, formado pelo mosto fermentado sem as leveduras, segue para as torres de destilação para obtenção do bioetanol. Já o leite de levedura, nome dado ao caldo de leveduras que sai da centrifugação, passa por um tratamento com ácido sulfúrico para reduzir o número de bactérias contaminantes (Oliveira Neto, 1990) e retorna à dorna, na qual será adicionada uma nova carga de mosto para um novo ciclo fermentativo (Meleiro and Maciel-Filho, 2000; Wheals et al., 1999). Este processo é repetido a cada ciclo fermentativo ao longo de toda a safra, que compreende em média nove meses. O reaproveitamento de leveduras é um procedimento economicamente importante para a usina. Isto porque a produção de um novo fermento iniciador com a quantidade de massa celular adequada necessita de tempo e um grande consumo de açúcar, o que torna o processo demorado, eleva os custos de produção e reduz a produtividade (Amorim et al., 2011; Meleiro and Maciel-Filho, 2000).

No processo industrial de fermentação, uma observação importante, sistematizada nos últimos anos, é que a população das linhagens de levedura não é homogênea e nem constante

durante a safra. As leveduras originais inoculadas em altas concentrações no início da safra muitas vezes são substituídas por linhagens selvagens em poucos dias durante o processo de fermentação (Basso et al., 2008; da Silva-Filho et al., 2005). Silva-Filho e colaboradores (2005) mostraram que, em alguns casos, essa substituição não era nem mesmo notada, pois não havia queda na produtividade da usina, possivelmente pelo fato das leveduras selvagens serem as melhores adaptadas às condições da usina.

Estas observações modificaram ao longo do tempo a estratégia de seleção de inóculos iniciais para a fermentação. Nas últimas duas décadas, linhagens nativas isoladas das próprias dornas vêm sendo selecionadas por possuírem características de interesse e boa permanência nas dornas (Basso et al., 2008). Estas linhagens passaram então a ser produzidas comercialmente em larga escala e utilizadas como inóculos iniciais ao invés ou em conjunto com linhagens de panificação.

Dentre as poucas cepas selecionadas com características desejáveis, destacam-se: BG-1, SA-1, CAT-1 e PE-2, utilizadas largamente na produção de bioetanol no Brasil. Na safra de 2007, CAT-1 e PE-2 foram utilizadas por cerca de 150 usinas, respondendo por 60% do bioetanol produzido no Brasil (Basso et al., 2008). Entretanto, essas linhagens vêm perdendo gradualmente seu desempenho ao longo dos anos (Luiz Carlos Basso, comunicação pessoal), principalmente devido às alterações operacionais do processo, como a redução da queima da cana-de-açúcar e o aumento do teor alcoólico da fermentação, e o aparecimento de linhagens selvagens mais robustas.

Os esforços de pesquisa no setor estão focados no sentido de aumentar o rendimento e a produtividade para que o bioetanol se torne ainda mais competitivo internacionalmente. Isto engloba desde o desenvolvimento de novas variedades de cana-de-açúcar, incluindo plantas transgênicas, como a expansão da área agrícola, inovações na linha de produção das usinas e seleção de linhagens de *Saccharomyces cerevisiae*. Apesar da importância biotecnológica, poucos estudos têm sido dedicados ao melhoramento genético das linhagens de levedura normalmente empregadas no processo fermentativo. Neste ponto, o escasso número de trabalhos sobre a compreensão genética e fisiológica das linhagens Brasileiras é um dos principais fatores limitantes para o progresso nesta área (Argueso and Pereira, 2010).

4. Melhoramento genético de leveduras

Com o avanço da biologia molecular e o desenvolvimento das técnicas de engenharia genética tornou-se possível a manipulação de genes de diversos organismos. Através da transformação de um organismo pela inserção ou modificação de genes e suas regiões regulatórias consegue-se hoje fazer com que o organismo expresse determinadas características que não são identificadas em populações naturais. Este procedimento visa principalmente à obtenção de características economicamente importantes, como por exemplo, o aprimoramento de processos biotecnológicos (Giga-Hama et al., 2007).

Ao longo dos anos, o melhoramento de linhagens de leveduras para uso industrial vem sendo foco de diversas pesquisas, uma vez que elas precisam apresentar características que permitam a sua sobrevivência em ambiente altamente estressante, como o encontrado nas dornas de fermentação. Dentre estas características destacam-se: estabilidade genética, alta tolerância e capacidade de produzir elevadas concentrações de etanol, osmotolerância (habilidade de fermentar soluções concentradas de carboidratos), alta viabilidade celular, tolerância à temperatura e tolerância a repetidos ciclos fermentativos (Dequin, 2001). Para isso, diversas técnicas de genética como cruzamento, transformação, fusão de protoplasto e mutagênese vêm sendo utilizadas.

Em um estudo com uma linhagem de levedura de laboratório, Alper e colaboradores (2006) obtiveram mutantes para o fator de transcrição *SPT15*, envolvido na maquinaria de transcrição global do organismo. Por estar na base do processo da transcrição, a mutação nesse gene, que codifica o TBP (*TATA-binding protein*), afeta um grande conjunto de genes. Os autores então selecionaram mutantes que apresentavam uma maior produção de etanol e observaram que estes isolados também possuíam uma maior resistência à molécula. Concluiu-se que a base dessa modificação era quantitativa, ou seja, um grande conjunto de genes estava envolvido nessas novas propriedades adquiridas pelas melhores linhagens. Isto demonstra que o entendimento da base fisiológica da produtividade exige a compreensão da expressão gênica global e a identificação de vias metabólicas alteradas, muito mais do que de genes específicos.

5. Maquinaria e técnicas moleculares para estudo de expressão gênica de *S. cerevisiae*

Transcrição gênica é o processo de síntese de RNA a partir de uma molécula de DNA molde codante. O início da transcrição ocorre quando a enzima RNA polimerase se liga a

região promotora de um determinado gene e inicia a síntese da molécula de RNA até atingir uma sequência terminadora, quando o processo é finalizado. Dessa forma, a molécula de RNA irá conter informações que serão posteriormente traduzidas para a síntese de uma proteína. Este processo ocorre em aproximadamente 75% do genoma de *S. cerevisiae* (Nagalakshmi et al., 2008), e até mesmo as regiões não codantes participam em importantes funções regulatórias (Berretta and Morillon, 2009).

A regulação da expressão gênica é controlada através de uma ação coordenada de múltiplos fatores de transcrição (elementos *trans*), os quais funcionam para ativar ou reprimir a transcrição através da ligação com os elementos de regulação (elementos *cis*) presentes na região promotora do gene. A transcrição de um gene pode ser constitutiva, de acordo com as condições fisiológicas e metabólicas que a levedura se encontra, ou induzida. Alguns genes são expressos apenas quando induzidos, por exemplo, na presença de um fator de estresse (mudanças de temperatura, redução de nutrientes, estresse oxidativo e osmótico, presença de agentes tóxicos, entre outros). Estes fatores de estresses são detectados por uma variedade de sistemas de sensores celulares que transmitem a informação para uma via de transdução de sinais. Os sinais são então transmitidos no final da via para os fatores de transcrição que se ligam ou são liberados dos elementos *cis*, permitindo a ativação ou repressão da transcrição de um gene específico. Finalmente, estas informações são identificadas pelos complexos de pré-iniciação da transcrição (TBP, TFIIA, TFIIB, TFIID, entre outros), que são responsáveis pelo recrutamento da enzima RNA polimerase e interação com co-ativadores e repressores que modulam a transcrição.

Informações sobre a quantidade e o momento do processo biológico que o gene é transcrito são de fundamental importância para estudos que visem à alteração da regulação de um determinado gene e a compreensão das respostas associadas a um determinado estresse. Neste ponto, ferramentas de genética molecular que identificam a expressão gênica global em diferentes tratamentos ou expostas a diferentes condições ambientais são amplamente utilizadas para esta finalidade, como também na busca de genes que possuam importantes aplicações biotecnológicas. Diversas metodologias podem ser utilizadas para avaliar a expressão gênica diferencial, como SAGE (*Serial Analysis of Gene Expression*) (Varela et al., 2005), microarranjos de DNA (Ehrenreich, 2006) e, mais recentemente, o sequenciamento de RNA (RNA-seq) por equipamentos de última geração (t Hoen et al., 2008; Marioni et al., 2008).

O microarranjo de DNA é uma ferramenta largamente empregada para a análise global de padrões de expressão gênica (Brown, 2002). Nesta técnica, os genes de um determinado

organismo estão depositados de maneira compartimentalizada em lâminas que posteriormente serão hibridados com o cDNA da amostra de interesse (Ehrenreich, 2006; Halasz et al., 2006), podendo assim comparar simultaneamente genes diferencialmente expressos em duas ou mais condições distintas. Os genes expressos em uma determinada condição (A) são previamente marcados com um fluoróforo (por exemplo, Cy5 – vermelho). A outra condição estudada (B) é então marcada com outro fluoróforo (Cy3 – verde). Os cDNAs marcados são então hibridados (separadamente ou em conjunto) com os genes presentes na lâmina. Finalmente, a intensidade da fluorescência formada para cada gene é proporcional à quantidade de RNA alvo presente em cada amostra comparada com a outra. Dessa forma, se determinado gene apresentar uma intensidade vermelha, podemos concluir que ele é mais expresso na condição A.

Pela técnica de RNA-seq (com variações estratégicas de metodologias), o RNA extraído do organismo é fragmentado e convertido por transcrição reversa em cDNA fita dupla utilizando-se iniciadores (*primers*) randômicos ou oligos dT (Nagalakshmi et al., 2008). O cDNA é então ligado a adaptadores, amplificado por PCR e submetido ao sequenciamento em modernos equipamentos (Illumina Solexa, Roche 454, SOLiD entre outros). A identidade das sequências é determinada pelo alinhamento com um genoma ou transcrito de referência, e o número de sequências referentes a um determinado gene é usado para quantificar sua expressão gênica, após as devidas normalizações intrínsecas da metodologia. Nesta técnica, o resultado é expresso em contagem absoluta de transcritos, enquanto que pela técnica de microarranjos, temos uma expressão relativa.

Devido à grande quantidade de dados gerados pelo sequenciamento das amostras, potentes computadores e a utilização de específicas ferramentas de bioinformática tornam-se essenciais para o processamento das informações (Costa et al., 2010). Além disso, a escolha dos parâmetros de análises constitui uma importante etapa e reflete na qualidade final dos dados obtidos (Oshlack et al., 2010). Ferramentas e metodologias de análises vêm sendo desenvolvidas constantemente, aprimorando dessa forma a utilização do RNA-seq como mecanismo de estudo de expressão gênica e possibilitando sua utilização para identificação de eventos de *splicing* alternativo, expressão de alelos específicos, edição de RNA e fusão de transcritos (Costa et al., 2010).

O alto desempenho das técnicas de microarranjos e RNA-seq deve-se ao fato da possibilidade de determinar a expressão diferencial de milhares de genes num único experimento. Entretanto, métodos de validação devem ser empregados em ambas as técnicas

a fim de identificar falso-positivos. Essa validação pode ser realizada utilizando *Northern blotting*, PCR em tempo real ou PCR semi-quantitativo (Fadiel and Naftolin, 2003). Aliado a isto, a característica de *S. cerevisiae* possuir um baixo controle pós-transcricional e a presença de íntrons em apenas 5% de seus genes (Juneau et al., 2007) torna a aplicação dos microarranjos de DNA e RNA-seq ainda mais eficientes, além de fornecer informações mais precisas sobre a relação entre a expressão gênica e as respostas fisiológicas observadas neste organismo.

Diversos estudos analisaram o perfil de expressão gênica de leveduras durante o processo fermentativo ou em resposta a algum tipo de estresse induzido. Alguns dos artigos sobre expressão gênica de leveduras são referentes à produção de sake (Hirasawa et al., 2007; Wu et al., 2006), vinho (Backhus et al., 2001; Erasmus et al., 2003; Marks et al., 2008; Rossignol et al., 2003; Varela et al., 2005; Zuzuarregui and Del Olmo, 2004), cerveja (Blieck et al., 2007; James et al., 2003; Olesen et al., 2002), produtos de panificação (Tanaka et al., 2006) e na produção de bioetanol a partir da sacarificação de amido (Li et al., 2010).

Em um estudo realizado com *Saccharomyces carlsbergensis*, responsável pela fermentação da cevada hidrolizada para a fabricação de cerveja, a técnica de microarranjos de DNA foi utilizada para avaliar a diferença de expressão entre 12 pontos durante o processo fermentativo (Olesen et al., 2002). Este resultado foi comparado com o trabalho de DeRisi e colaboradores (1997), que avaliaram a expressão gênica com ensaios de microarranjos realizados em laboratório. Os resultados mostram similaridades, mas também diferenças significativas na comparação entre as duas situações, principalmente para fatores de estresse relacionados com o processo de fermentação em uma escala industrial. Este experimento foi importante, pois forneceu inúmeras informações com relação à regulação geral da expressão durante o processo fermentativo e também informações sobre o perfil de expressão gênica individual e em grupo. Este amplo entendimento fisiológico vem fornecendo diversas informações para o desenvolvimento de novas linhagens melhoradas de leveduras através da engenharia genética (Olesen et al., 2002).

6. Genética de linhagens de *S. cerevisiae* Brasileiras

Apesar de terem uso amplamente difundido na indústria Brasileira de bioetanol, as cepas isoladas a partir de culturas mistas de contaminantes selvagens ainda permanecem pouco conhecidas do ponto de vista genético. Essa falta de informações básicas sobre a biologia dessas linhagens representa uma barreira para o eventual melhoramento genético e

melhor aproveitamento de seu potencial. Apenas recentemente as primeiras informações sobre as principais linhagens utilizadas atualmente no Brasil (CAT-1 e PE-2) foram reportadas (Argueso et al., 2009; Babrzadeh et al., 2012; Stambuk et al., 2009). Estes estudos demonstraram que o genoma dessas linhagens é organizado de tal maneira a tolerar as condições de estresse impostas no processo de produção de bioetanol, principalmente através da habilidade que possuem de variar o número de cópias de genes específicos de acordo com as condições do meio. Estas características conferem assim uma vantagem adaptativa que pode ser observada diretamente pelos seus excelentes desempenhos fermentativos.

Ainda neste ponto, nenhum estudo relata genes diferencialmente expressos em linhagens Brasileiras durante o processo fermentativo industrial para a produção de bioetanol, dificultando a escolha de fatores para o melhoramento genético nesta condição específica. Com isso, fica claro que qualquer trabalho a ser desenvolvido com o intuito de se alterar a expressão de um determinado gene em leveduras utilizadas em processos industriais deve ser precedido de um estudo detalhado da expressão deste gene nas condições das usinas e não apenas em ensaios de laboratório. No caso da produção de bioetanol no Brasil, a manipulação genética das linhagens utilizadas no processo deve ser realizada de maneira minimamente invasiva, preservando-se assim as características de robustez e ótimo desempenho fermentativo (Argueso and Pereira, 2010). Esta manipulação só é obtida com êxito quando se conhece em detalhes a organização do genoma da linhagem a ser modificada e suas informações genéticas desencadeadas na condição de interesse.

OBJETIVOS

O conhecimento da estrutura e organização genômica, assim como os mecanismos que o controlam são de fundamental importância para o sucesso de programas de melhoramento genético de linhagens de *Saccharomyces cerevisiae* para produção de bioetanol. Assim, o presente trabalho teve como principal objetivo a avaliação do genoma da linhagem PE-2 e a caracterização de seu programa gênico desencadeado ao longo da fermentação industrial. Os objetivos específicos compreenderam o desenvolvimento de metodologias para manipulação genética de PE-2 e a identificação genotípica de linhagens de *S. cerevisiae*.

APRESENTAÇÃO DO TRABALHO

Este trabalho objetivou o estudo de caracterização genômica, além da análise de expressão e manipulação gênica da linhagem de *Saccharomyces cerevisiae* PE-2, utilizada na produção de bioetanol no Brasil. O trabalho foi dividido em cinco capítulos, organizado cronologicamente de acordo com a obtenção dos dados.

Capítulo 1. A ausência de informações genéticas sobre uma das principais linhagens de *S. cerevisiae* utilizada no Brasil para produção de bioetanol (PE-2) motivou nosso grupo a sequenciar o genoma de um isolado haplóide derivado (JAY291). Este foi o primeiro genoma sequenciado de uma linhagem Brasileira, que contou também com a caracterização de sua estrutura diplóide. O estudo revelou a complexidade do genoma de PE-2, caracterizado por um elevado grau de heterozigosidade, alta frequência de rearranjos cromossômicos e a presença de genes ausentes na linhagem padrão de laboratório. Este genoma foi a base para a concepção de todos os estudos apresentados posteriormente neste trabalho.

Capítulo 2. A cariotipagem (PFGE) é a principal técnica de identificação de leveduras no processo industrial de produção de bioetanol. Entretanto, dispense de um longo tempo para preparo e análise das amostras. A partir dos dados do genoma de PE-2, foram desenvolvidos onze marcadores moleculares para identificação de linhagens de *S. cerevisiae*. Estes marcadores contêm regiões polimórficas localizadas em sequências codantes, e apresentaram um alto poder discriminatório, tanto para linhagens com mesma origem tecnológica quanto geográfica. Esta tecnologia foi recentemente licenciada e vem sendo utilizada comercialmente para genotipagem de amostras provenientes de usinas, com uma eficiência superior a obtida por PFGE.

Capítulo 3. A auxotrofia é uma estratégia largamente utilizada em ensaios de manipulação genética de *S. cerevisiae* para diferenciação das células transformadas. Uma vez que as linhagens utilizadas no processo industrial de produção de bioetanol são selvagens, estas não possuem marcas de auxotrofia, e as manipulações são realizadas principalmente com genes de resistência a antibiótico. Assim, este estudo objetivou o desenvolvimento de um isolado diplóide de PE-2 auxotrófico para uracila (*ura3*). As manipulações genéticas foram realizadas de

maneira eficiente sem que as desejáveis características fermentativas da linhagem fossem perdidas durante as etapas de transformação. Esta linhagem “domesticada” deverá ser utilizada como uma plataforma biológica em estudos posteriores visando o melhoramento genético de PE-2.

Capítulo 4. O conhecimento do programa de expressão gênica das leveduras ativado ao longo da fermentação é de fundamental importância para os ensaios de manipulação genética. Esses dados são ainda mais relevantes quando obtidos em condições semelhantes às encontradas no processo industrial. Neste capítulo é apresentado um estudo de expressão gênica de PE-2 ao longo de duas fermentações em escala industrial: sob uma condição típica de usina e outra sob efeito de floculação devido à ação bacteriana. Os dados obtidos nestas análises revelam de forma inédita a programação gênica de PE-2 ao longo da fermentação e os mecanismos utilizados contra os estresses impostos pelo processo. As informações apresentadas possuem um grande potencial biotecnológico e deverão ser utilizadas posteriormente na manipulação de PE-2 e linhagens relacionadas.

Capítulo 5. Apesar das qualidades fermentativas das linhagens comerciais disponíveis no Brasil para produção de bioetanol, estas vêm perdendo gradualmente o desempenho ao longo dos últimos anos. O fato de serem selvagens abre perspectivas para que modificações pontuais possam ser feitas visando um incremento de desempenho na fermentação. A partir de dados de expressão gênica de PE-2 obtidos em escala industrial, as regiões promotoras de alguns genes foram substituídas de maneira a elevar sua expressão e torna-los constantemente expressos ao longo de todo o processo. Com isto, foi possível o desenvolvimento de linhagens com características fermentativas superiores ao isolado selvagem. Novos ensaios estão sendo realizados para validação dos dados. Entretanto, nosso grupo acredita que os resultados aqui apresentados confirmam a viabilidade da estratégia para o desenvolvimento de uma segunda geração de linhagens ainda mais eficientes para a produção de bioetanol.

CAPÍTULO 1

Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production.

Juan Lucas Argueso ^{1,9,10}, Marcelo F. Carazzolle ^{3,9}, Piotr A. Mieczkowski ^{6,9}, Fabiana M. Duarte ³, Osmar V. Carvalho-Netto ³, Silvia K. Missawa ³, Felipe Galzerani ³, Gustavo G. L. Costa ³, Ramon O. Vidal ³, Melline F. Noronha ³, Margaret Dominska ¹, Maria G. S. Andrietta ⁴, Sílvia R. Andrietta ⁴, Anderson F. Cunha ⁵, Luiz H. Gomes ⁷, Flavio C. A. Tavares ⁷, André R. Alcarde ⁸; Fred S. Dietrich ^{1,2}, John H. McCusker ¹, Thomas D. Petes ¹, and Gonçalo A. G. Pereira ^{3,10}

¹ Department of Molecular Genetics and Microbiology; and ² Institute for Genome Sciences and Policy, Duke University Medical Center, Durham-NC, USA. ³ Laboratório de Genômica e Expressão, Departamento de Genética e Evolução, Instituto de Biologia; and ⁴ Laboratório de Biotecnologia e Bioprocessos, Centro Pluridisciplinar de Pesquisas Químicas e Biológicas, Universidade Estadual de Campinas, Campinas-SP, Brazil. ⁵ Departamento de Genética e Evolução, CCBS, Universidade Federal de São Carlos, São Carlos-SP, Brazil. ⁶ Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill-NC, USA. ⁷ Departamento de Genética; and ⁸ Departamento de Agroindústria, Alimentos e Nutrição, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba-SP, Brazil. ⁹ These authors contributed equally to this work. ¹⁰ Corresponding authors.

Corresponding authors:

Juan Lucas Argueso, PhD

Department of Molecular Genetics and Microbiology, Duke University Medical Center,
371 CARL Building, Box 3054, Durham-NC, 27710, USA.

Phone: 1-919-684-5814; Fax: 1-919-684-6033; email: lucas.argueso@duke.edu

Gonçalo A. G. Pereira, PhD

Laboratório de Genômica e Expressão, Departamento de Genética e Evolução, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas-SP, 13083-970, Brazil.

Phone: 55-19-3521-6237; Fax: 55-19-3521-6235; email: goncalo@unicamp.br

ABSTRACT

In the search for renewable sources of energy, bioethanol stands out as a benchmark biofuel because its production is based on a proven technological platform. Bioethanol is produced mainly from the fermentation of carbohydrates derived from agricultural feedstocks by the yeast *Saccharomyces cerevisiae*. One of the most widely adopted strains is PE-2, a heterothallic diploid naturally adapted to the sugar cane fermentation process used in Brazil. Here we report the molecular genetic analysis of a PE-2 derived diploid (JAY270), and the complete genome sequence of a haploid derivative (JAY291). The JAY270 genome is highly heterozygous (~2 SNPs per kilobase), and has several structural polymorphisms between homologous chromosomes. These chromosomal rearrangements are confined to the peripheral regions of the chromosomes, and appear to reflect ectopic homologous recombination between repetitive DNA sequences. Despite the complex karyotype of JAY270, this diploid, when sporulated, had a high frequency of viable spores (93%). Crosses of haploids derived from JAY270 to a haploid of the unrelated laboratory strain S288c also resulted in diploids that had good spore viability (75-95%). Thus, the rearrangements that exist near the ends of chromosomes do not impair meiosis and spore viability, as they do not span regions that contain essential genes. This observation is consistent with a model in which the peripheral regions of chromosomes represent plastic domains of the genome that are free to recombine ectopically and experiment with alternative structures. We also explore features of the JAY270 and JAY291 genomes that help explain their high adaptation to industrial environments, exhibiting desirable phenotypes such as high cell mass production and fermentation kinetics, high temperature growth and oxidative stress tolerance. The genomic manipulation of such strains could enable the creation a new generation of industrial organisms, ideally suited for use as delivery vehicles for future bioenergy technologies.

INTRODUCTION

As concerns mount over the alarming effects of climate change and of the continued volatility in petroleum markets, nations throughout the world are increasingly adopting policies to promote the use of renewable and domestic sources of energy (Robertson et al. 2008). Among the most viable alternatives, bioethanol stands out as a benchmark biofuel because its production is based on a proven technological platform. Bioethanol is produced through microbial fermentation of carbohydrates derived from agricultural feedstocks, mainly starch and sucrose. While the United States and Brazil are the dominant players in bioethanol, sharing approximately 70% of the global market, their production systems differ in many respects (Sanderson 2006; Goldemberg 2007). North American bioethanol is currently produced from enzymatically-hydrolyzed starch from grains. Farming this type of feedstock is energy intensive, resulting in a final energy balance that is only marginally positive. This process is currently regarded as a transitional technology, which in the coming years is expected to give way to advanced cellulosic biofuels (Gura 2009). The Brazilian system uses sugar cane as feedstock, a tropical grass crop that abundantly accumulates sucrose that can be converted to bioethanol directly by the yeast *Saccharomyces cerevisiae* without enzymatic pre-treatment. Among other advantages, sugar cane is semi-perennial and has symbiotic interactions with nitrogen-fixing microorganisms, allowing this system to produce an energy surplus estimated at about eight-fold (Goldemberg 2007; Robertson et al. 2008).

In the microbial fermentation process used in Brazil, the yeast inoculum is often recycled from one fermentation tank to the next (Wheals et al. 1999). In many distilleries, this recycling spans the entire eight-month sugar cane harvesting season (Basso et al. 2008), imposing both biotic and abiotic stresses on the fermenting yeast strain. While yeast strains genetically bred for high-efficiency in fermentation have been available since the 1980's (Tavares and Echeverrigaray 1987), they are often outcompeted in the industrial environment by more robust "wild" yeasts that contaminate the sugar cane stalks (Wheals et al. 1999; da Silva et al. 2005). In the 1990's, an alternative strategy was adopted: selection among the wild yeast contaminants for those that combined high fermentation efficiency with prolonged persistence in the system. In recent years, such strains have been widely adopted by the industry. One of the most successful examples is the PE-2 wild isolate (Basso et al. 2008), currently used by ~30% of Brazilian distilleries, generating about 10% of the world's bioethanol supply.

Thirteen years ago, the *S. cerevisiae* laboratory strain S288c became the first eukaryote to have its genome completely sequenced (Goffeau et al. 1996). Since then, other haploid

strains from diverse backgrounds have been sequenced (RM11-1a, YJM789, M22, YPS163, and AWRI1631) (*S. cerevisiae* RM11-1a Sequencing Project - <http://www.broad.mit.edu/>) (Wei et al. 2007; Borneman et al. 2008; Doniger et al. 2008), and more recently a large scale effort to determine the genome sequences of many others has been completed (Liti et al. 2009).

Extensive analysis has been done to examine the nucleotide sequence diversity between these strains (Liti et al. 2009; Schacherer et al. 2009), whereas studies of structural variation have mostly focused on comparisons of *S. cerevisiae* to other related species (Fischer et al. 2000; Goffeau 2004; Scannell et al. 2007; Gordon et al. 2009). Microarray-based whole genome hybridization studies of wild, industrial, and laboratory *S. cerevisiae* strains (Winzeler et al. 2003; Carreto et al. 2008; Schacherer et al. 2009; Faddah et al. 2009) have uncovered a recurrent pattern of copy number variation (CNV) near the ends of chromosomes, suggesting a role for repetitive DNA sequences in structural genome diversification. Despite these valuable insights, two central questions regarding the role of chromosomal rearrangements in genome evolution in *S. cerevisiae* remain unanswered: First, it is still unclear how these rearrangements contribute to long-term fitness in natural environments; and second, it is not known if they are compatible with the formation of viable meiotic spores that would allow their sexual dissemination between natural *S. cerevisiae* populations.

In this study, we report a detailed molecular genetic characterization of a PE-2 derived diploid (JAY270), and the complete genome sequence of a derived haploid (JAY291). We have found extensive structural differences between the genome of JAY270 and those of other sequenced *S. cerevisiae* strains, including chromosomal translocations and insertions of large blocks of DNA absent in the reference strain's genome. These differences were confined to the peripheral regions of the chromosomes, while the central portions have remained structurally conserved. This pattern of structural variation was fully compatible with the formation of viable meiotic spores, and in some cases, the chromosome rearrangements resulted in the amplification of genes implicated in environmental stress tolerance. This observation supports a model in which *S. cerevisiae* chromosomes are organized in two structural domains (Pryde et al. 1997). Our data suggest a general chromosome organization in which the central core sectors, harboring the essential genes, are refractory to rearrangements, while the peripheral regions are highly plastic and are free to undergo ectopic recombination.

RESULTS

Genetic and phenotypic characterization

Before initiating a detailed characterization of PE-2, we used Pulse-Field Gel Electrophoresis (PFGE) to show that the commercially-available stocks of this strain contain a mixture of cells that have slightly diverged karyotypes (Supplemental Fig. 1). Consequently, we purified a single colony isolate with a representative karyotype. This isolate, JAY270, was indistinguishable from the original PE-2 culture in phenotypic tests (Fig. 1a, and data not shown) and was used in all subsequent experiments.

Upon induction of meiosis, JAY270 displayed rapid and efficient sporulation producing mostly asci with four ascospores. We dissected 104 tetrads and observed four viable spores per tetrad in 88 (93.3% spore viability). By PCR analysis, we found 2:2 segregation at the mating type locus (*MAT*; data not shown). These results show that JAY270 is a diploid, does not contain recessive lethal mutations, and is not heterozygous for chromosome aberrations that span essential genes. The haploid segregants derived from JAY270 had stable *MATa* or *MAT α* genotypes, indicating that the strain is naturally heterothallic. We crossed haploid strains (both *MATa* and *MAT α*) derived from JAY270 to S288c-derived strains of the opposite mating type. The hybrid diploids produced in these crosses sporulated normally, had high spore viability (75-95%), and exhibited Mendelian segregation for all auxotrophic markers from the S288c-derived parent. These results argue that the chromosome structure of JAY270 is mostly collinear with that of S288c, at least in the regions required for viability.

We compared JAY270 and its derivatives to S288c in several phenotypic tests. In sugar cane extract fermentation assays that simulate the conditions found in the industry (Fig. 1a), JAY270 and all four spore clones from a JAY270 tetrad (segregants JAY289, JAY290, JAY291 and JAY292), produced roughly 50% more ethanol than the S288c strain. The observation that all four spore clones were equally proficient in ethanol production argues that the major alleles contributing to this complex phenotype are homozygous in the JAY270 diploid. We also compared growth phenotypes under a variety of stress conditions. Tolerance to ethanol, acid washing and short-term heat killing were not differently affected in these strains (Supplemental Fig. 2); however, JAY270 was much more tolerant to oxidative stress and high temperature growth than S288c (Fig 1b). When grown in the presence of menadione, an inducer of reactive oxygen species, JAY270 grew normally at concentrations that were inhibitory to S288c. JAY270 was also highly tolerant to temperature stress. While S288c had reduced growth rates at 37C, and did not form isolated colonies at 41C, JAY270 formed larger colonies at 37C than it did at

30C (Supplemental Fig. 3). At 41C, JAY270 displayed slower growth, but maintained full viability. These traits are likely important for long-term viability and competitiveness in the industrial setting, where oxidative stress (Landolfo et al. 2008) and heat are generated during high-pitch fermentation (fermentation involving large inocula).

Finally, we examined the patterns of genetic inheritance associated with two of the most important industrial traits of JAY270: cell mass accumulation and kinetics of ethanol production. To bypass the influence of factors specific to sugar cane extract (*e.g.* sucrose utilization), the fermentations were carried out in rich media with 10% glucose. We mated an S288c-isogenic strain to a haploid spore derived from JAY270 (JAY291) and generated hybrid diploids. Two reciprocal F₁ generation diploids (JAY361 and JAY365) were created by mating ρ^0 mitochondrial *petites* derived from either parent (Fox et al. 1991). These two diploids were phenotypically indistinguishable from each other (data not shown), indicating that differences between the mitochondrial genomes of S288c and JAY291 do not have a major influence on the traits examined. We then evaluated the JAY291 and S288c haploid parents and the F₁ hybrid JAY361 in time-course fermentations. JAY291 displayed a higher growth rate and finished the fermentation with a cell density 34% higher than S288c; the JAY361 hybrid had an intermediate phenotype (Fig. 2a). Under the conditions used in this assay all three strains reached the same final concentration of ethanol at the end of the fermentation (Fig. 2b), but, JAY291 and JAY361 were much more efficient, displaying higher kinetics of ethanol production and requiring ~30% less time than S288c to complete the fermentation. We have not tested diploidized versions of the JAY291 and S288c parents in these assays, therefore we cannot firmly ascertain dominance effects associated with these traits. However, if ploidy is not an influencing factor, our results would suggest that the cell mass accumulation trait is semi-dominant, whereas the kinetics of ethanol production trait is fully dominant.

We sporulated and dissected asci from JAY361 to isolate F₂ haploids segregating genomic segments from the two parent strains. Individual clones in this progeny were then examined for final cell mass accumulation, and ethanol concentration after 8 hours of fermentation (the point of highest differential between the parents) as an indicator of ethanol production kinetics. The phenotypic distributions for these two traits in the progeny were quite distinct, and gave preliminary indications of the genetic control associated with each. The distribution of cell mass accumulation (Fig. 2c) appeared to be bimodal, with two distinct peaks centered around the phenotypes of the parents and a valley which corresponded roughly to the phenotype of the F₁ hybrid. One interpretation of this pattern is that it was created by

segregation of the two alleles of a single gene that is the major determinant of the cell mass accumulation phenotype. In addition to this major locus, the segregation of other minor contributing alleles was responsible for the bell-shaped distribution around each peak. The distribution of ethanol production kinetics (Fig. 2d) was more typical of a quantitative trait controlled by loci with roughly equal contributions to the phenotype. From this limited initial analysis, we can estimate that three or four genes are responsible for the segregation pattern observed for ethanol production kinetics.

We also asked whether there is a correlation between cell mass accumulation and ethanol production kinetics. This analysis (Fig. 2e) suggested that the two traits are largely independent of each other, and are presumably controlled by different genes. The calculated broad-sense heritabilities associated with both traits were very high ($H^2 > 0.90$), indicating that most of the phenotypic variance observed between F_2 individuals was due to differences in genetic composition.

Molecular karyotype analysis

We compared the molecular karyotypes of JAY270 and S288c using a combination of PFGE, and microarray-based Comparative Genomic Hybridization (CGH-array) and Band-array (Argueso et al. 2008). PFGE analysis revealed clear chromosome (Chr) length polymorphisms between S288c and JAY270 (Fig. 3a). This analysis also showed polymorphisms between several pairs of homologs within the JAY270 diploid. Image tracing of the separated chromosomes in JAY270 revealed that some chromosomes were represented by two bands, each present in one copy in the diploid. For example, Chr6 is represented by short (Chr6S) and long (Chr6L) homologs, differing in size by about 60 kb. Interestingly, karyotypic analysis of the meiotic haploid spores revealed parental-sized chromosomes in JAY289 and JAY290, and chromosomes with the sizes expected for reciprocal recombinants between Chr6S and Chr6L in JAY291 and JAY292 (Fig. 4a; second band from the bottom). In addition to Chr6, at least Chr3, Chr5, Chr9, Chr11, Chr14, and either Chr7 or Chr15 had length polymorphisms between homologs in JAY270.

We examined the structural polymorphisms between the Chr6 homologs by excising their respective chromosomal bands from PFGE, labeling the DNA and hybridizing them to microarrays covering the entire S288c reference genome (Band-array; Fig. 5). The JAY270 Chr6 homologs differed from each other and from Chr6 in S288c only near the chromosome ends, in regions distal to the first and last essential genes (*SEC53* and *RPN12*). This analysis

showed that Chr6L contained a ~16 kb region on the left arm that was missing in Chr6S (including the *AGP3* gene), and that Chr6S contained a ~12 kb translocation from the right arm of Chr1 (including the *YAR064W* gene) that was missing in Chr6L. Southern analysis (Fig. 4b-c) confirmed this arrangement and revealed an additional translocation involving *AGP3* on one of two Chr10 homologs of JAY270. Notably, the cumulative Band-array hybridization signal for Chr6L (~322 kb in size, Fig. 2a), accounted for only ~250 kb of sequences, suggesting the presence of about 70 kb of DNA that did not hybridize to the microarrays.

We used CGH-arrays to look for sequences that were missing or duplicated in JAY270 relative to S288c (CNVs; Fig 3b). Most regions of the genome were present in equal dose between these strains. The regions shown in green near the telomeres of Chr1, Chr4, Chr6, Chr10, and Chr15 were underrepresented in the JAY270 genome. These could be either missing in JAY270, or represent diverged sequences that do not hybridize to the S288c-based microarrays. The underrepresented regions that were not telomeric corresponded to regions containing tandemly repeated genes in S288c (*HXT7*, *HXT6*, *HXT3* and *ENA5*, *ENA2*, *ENA1* in Chr4; and *ASP3-1* to *ASP3-4* in Chr12) that likely contracted in JAY270 or expanded in S288c as a consequence of unequal crossing-over. The regions shown in red represent sequences that were amplified in JAY270 relative to S288c. An amplification signal detected in the central section of Chr15 corresponded to the *HIS3* marker locus. This gene is present in JAY270, but was intentionally deleted in the derivative of S288c that was used in the CGH-array experiments. Thus, we can effectively detect a single amplified gene by these methods.

The amplification at the left end of Chr16 spans the *SAM3* and *SAM4* genes that are involved in metabolism of S-adenosylmethionine (AdoMet), a key biochemical cofactor that participates in a variety of metabolic pathways. The Sam4p has been proposed to participate in the recycling of AdoMet from the inactive, and possibly toxic, (*R,S*)-AdoMet isomeric state back to the biologically essential form (*S,S*)-AdoMet (Vinci and Clarke 2007). Based on CGH-array and Southern blot data (Fig. 3b, and data not shown), we estimate that four copies of these genes are present in JAY270 compared to two in the laboratory diploid strain.

Another interesting gene amplification resulted in two peaks at the left ends of Chr6 and Chr14. In S288c these regions contain, respectively, the duplicated *SNO3* and *SNZ3*, and *SNO2* and *SNZ2* genes which are involved in vitamin B6 metabolism, and have a role in oxidative stress tolerance (Padilla et al. 1998; Ehrenshaft et al. 1999; Rodriguez-Navarro et al. 2002). To determine the extent and chromosomal distribution of the *SNO/SNZ* gene amplifications in JAY270, we did a Southern analysis of pulse-field gels for this region (Fig. 4d).

We detected nine copies of the *SNO/SNZ* genes in JAY270, compared to four in the S288c diploid. Consistent with this observation, we found that expression levels of these genes were up-regulated about four-fold in JAY270 relative to S288c (Supplemental Table 1). These results also showed that *SNO/SNZ* genes were redistributed among the chromosomes in JAY270 with no copies on either Chr6 homolog and new copies on Chr9, Chr10, Chr13 or Chr16, Chr7 or Chr15, and Chr4 (Fig. 4d). Strikingly, three of the extra copies were hemizygous (only one copy of a gene is present in a pair of homologous chromosomes) in the JAY270 diploid as could be inferred from their Mendelian segregation among the haploid spores JAY289-JAY292.

Genome sequencing of JAY291, a haploid spore derived of JAY270

In addition to the molecular genetic analysis described above, we employed massively parallel DNA sequencing to characterize the genome of JAY270. To avoid complications due to heterozygosity in the diploid, we sequenced the genome of a haploid derivative, JAY291. Two independent DNA sequencing platforms were used, Roche/454 and Illumina, resulting in 162-fold coverage (Table 1). Despite this high level of redundancy, because both of these methods generated relatively short sequence reads (<250 bp), we were still unable to assemble the genome into one single contiguous sequence per chromosome. Instead, our *de novo* assembly of the JAY291 genome resulted 452 genomic fragments (contigs; Materials and Methods). As expected, when we compared these contigs to the S288c reference genome, it became clear that most interruptions occurred at regions of dispersed repetitive DNA sequences, mostly Ty retrotransposon insertions. As described below, we compared the sequences of JAY291 to those of other sequenced *S. cerevisiae* strains: S288c, RM11-1a, and YJM789 (Wei et al. 2007). These sequence comparisons were done to characterize chromosome structure differences between these strains, as well as to determine the level of nucleotide sequence divergence.

Chromosome structure polymorphisms

From the PFGE profiles of JAY270 and its haploid descendants (Figs. 3a and 4a) it was clear that this diploid has chromosomes that vary in size, not only in comparison to S288c, but also between the two homologs. To investigate the nature of these rearrangements, we focused our analysis on Chr6. In JAY270, the two homologs were approximately 261 kb (Chr6S) and 322 kb (Chr6L) in size. The Chr6 of JAY291 is a recombinant between Chr6S and Chr6L, and is about 290 kb in size. We assembled the complete sequence of Chr6 in JAY291 by searching the 452 whole genome contigs for known Chr6 sequences from the S288c, RM11-1a and

YJM789 strains; this search identified 14 candidate contigs. We then used PCR to amplify across the small gaps between the contigs and to position telomeric repeats, confirming their relative order and orientation (data not shown). The size of the resulting assembled sequences (including one X and one Y' element as each telomere) was 290 kb, matching the observed PFGE size. The chromosome structure and nucleotide sequence in the central portion of Chr6 were very similar to those of S288c, whereas in the regions near the telomeres the sequences were more diverged or completely different (Fig. 6a). At the left end of the Chr6 alignment, we found two segments (9.1 and 19.3 kb) that are absent in the S288c genome. We also aligned the JAY291 chromosome to the Chr6 sequence from *Saccharomyces paradoxus* (Kellis et al. 2003), the closest known relative of *S. cerevisiae*, and observed continual synteny through most of Chr6, including the entire left arm. This alignment suggests that the 9.1 and 19.3 kb insertions were present in the last common ancestor of *S. cerevisiae* and *S. paradoxus*, and that they were probably lost from the progenitor of S288c. The only structural difference in the central section of Chr6 was a Ty2 retrotransposon element insertion which is present in S288c (*YFLWTy2-1*), but absent in JAY291. Chr6 had no full-length retrotransposons in JAY270, but Ty elements were found in all other chromosomes (Supplemental Fig. 4).

Although we concentrated our sequencing efforts on the JAY291 haploid, from our analysis of the sequences of S288c, RM11-1a, and YJM789 and from our examination of the microarray data and Southern analysis (Figs. 3-5), we were able to predict the structures at the left ends of the two Chr6 homologs in the JAY270 diploid. We then designed specific PCR primers to confirm our predictions (Supplemental Fig. 5). Repetitive DNA sequences were found at the breakpoints (shown as boxed letters in Fig. 6b) of all chromosomal rearrangements in this region, suggesting that they formed through ectopic homologous recombination.

We determined that JAY270 Chr6S has a portion of Chr1 that includes the *YAR064W* gene attached to the left end of Chr6 (Fig. 4c). This translocation is also present in RM11-1a and YJM789, and the breakpoint is between *YAR062W* in Chr1 (boxed "A") and *YFL051C* in Chr6, both are pseudogene members of the *FLO* gene family. The *YFL051C* pseudogene is also the site of the 19.3 kb region which is present in JAY270 Chr6L and deleted in Chr6 from S288c. While *YFL051C* is short and non-functional in S288c, in JAY270 Chr6L this ORF extends into the 19.3 kb region to form a full length *FLO* gene (boxed "B"). Near the left end of this insertion region, there was a 4.3 kb translocated section of S288c Chr10, with breakpoints at *HXT8* (boxed "C"; a hexose transporter gene family member) and *YJL216C* (boxed "D"; a gene with similarity to maltase genes). Finally, the leftmost rearrangement in JAY270 Chr6L contains

the 9.1 kb sequence not represented in S288c. The breakpoint for this rearrangement was at the *THI5* gene (boxed “E”; a member of a gene family involved in the thiamine biosynthesis). In S288c, the *SNO3* and *SNZ3* genes are located immediately distal to *THI5*, whereas this region of Chr6L of JAY270 contains a copy of the *MPR2* gene involved in tolerance to oxidative stress and ethanol toxicity (Du and Takagi 2007). *MPR2* is also present in the RM11-1a and Σ 1278b strains, but strikingly, the genes in *MPR* family are found at different genomic locations in these strains: in RM11-1a, *MPR2* is at the left end of Chr10; whereas in Σ 1278b, *MPR2* is at the right end of Chr10 and *MPR1* is at the left end of Chr14 (Takagi et al. 2000).

This detailed analysis also allowed us to examine the rearrangements between the two Chr10 homologs in JAY270 (arbitrarily designated A and B). The left end structure of Chr6L was essentially identical to that found in Chr10 of RM11-1a and in JAY270 Chr10A (inherited by the JAY289 and JAY292 spores). A similar Chr6/Chr10 translocation with a breakpoint at *YJL216C* (boxed “D”) has been described recently in the M22 vineyard strain (Doniger et al. 2008). The Chr10 structure in YJM789 is similar to RM11-1a, M22, JAY270 Chr10A, with the exception of the extreme left end at positions distal to *THI5* (boxed “E”) where YJM789 is similar to S288c Chr6. Interestingly, the YJM789 Chr10 configuration is precisely what would be predicted from a crossover occurring between S288c Chr6 and RM11-1a Chr10 anywhere between *THI5* and *YFL051C*. An analogous meiotic crossover event within translocated chromosomal segments has been recently observed in the S288c/Y101 laboratory hybrid (Faddah et al. 2009). The second JAY270 Chr10 homolog (Chr10B; inherited by JAY290 and JAY291) had at its left end a ~15 kb translocated segment not found in the S288c genome. The breakpoint in this chromosome is near an unannotated Ty1 delta LTR element (boxed “F”) that is located on S288c Chr10 between SGD coordinates 29000 and 29500 and is also present in the JAY291 Chr10 sequence. Finally, it is important to note that all rearrangements described above occurred at sites distal to the first essential genes in the Chr6 (*SEC53*) and Chr10 (*PRP21*).

Nucleotide polymorphisms

We compared the sequences of JAY291 to S288c, estimating the number of SNPs at 5.4 per kilobase (about 65000 for the entire genome). To compare JAY291 to other *S. cerevisiae* strains, we analyzed a ~49 kb region from Chr14 that has been sequenced for several unrelated strains (Steinmetz et al. 2002) and used it to construct an unrooted phylogenetic tree (Fig. 7a). This analysis indicated that JAY291 is highly diverged relative to S288c, YJM789, and RM11-1a and, therefore, will be useful in characterizing the rich sequence diversity present in *S.*

cerevisiae.

We also sought to estimate the degree of heterozygosity in the JAY270 diploid genome. We performed a limited sequence analysis of JAY292, a haploid spore derived from the same tetrad as JAY291. Because the genomic coverage in this case was relatively low (15-fold with Illumina paired-end reads), we did not attempt to assemble the genome *de novo*. Instead, we assembled the JAY292 sequence using the JAY291 contigs as reference. After filtering out repetitive elements, regions of high divergence, and regions of low quality and/or coverage, we estimated a density of about 1.3 SNPs/kb in the single-copy genomic regions. Because JAY291 and JAY292 are sibling haploid strains, about half of their genomes should be identical by descent and we found that about half of the genomic regions had a very low level of polymorphisms (data not shown). We estimate, therefore, that the JAY270 diploid has about 2.6 SNPs/kb between allelic regions in homologous chromosomes.

We also examined heterozygosity restricting the analysis to the Chr6 sequences of JAY291 and JAY292. Because these two haploids carry the two reciprocal products of a meiotic crossover between the two parental Chr6 homologs (Fig. 4a), it should be possible to directly find all Chr6 SNPs in the diploid by comparing the JAY291 and JAY292 sibling sequences (assuming that they formed as a result of exactly one crossover). We found 465 SNPs in the 208 kb region common to both the Chr6L and Chr6S homologs (Fig. 5). The Chr6 SNP density (~2.2 SNPs/kb) was consistent with our estimate for the whole genome.

Overview of specific gene polymorphisms important for bioethanol production

We compared the sequences of several JAY291 genes to those of other strains to establish possible links between the genotype and the observed phenotypes. High-Temperature Growth (HTG) has been investigated in *S. cerevisiae* as a model for complex quantitative inheritance (Steinmetz et al. 2002; Sinha et al. 2008), and three major contributing loci have been identified: *NCS2*, *MKT1*, and *END3*. We found that JAY291 had all three of the alleles that are related to HTG⁺ (Fig. 7b). However, HTG, as is often the case for quantitative traits, is known to be influenced by genetic background (Sinha et al. 2006). We therefore tested the role of these polymorphisms in the JAY291 background by constructing isogenic strains carrying HTG⁻ allele replacement versions of *NCS2* and *MKT1* from S288c, and *END3* from YJM789. We found that the allele replacement strains, individually, were not compromised for HTG, at least within the resolution of our co-culture competition assays (~5 fold; data not shown). This result was analogous to the background-specific phenotypic response described by Sinha et al.

(2006), and suggested that other uncharacterized alleles present in the JAY291 genome are contributing to its distinct HTG⁺ phenotype.

We also examined the sequence of the *HO* gene that encodes the endonuclease that stimulates mating type switching in homothallic strains (Herskowitz 1988). The *ho* allele in JAY291 has three missense mutations, including the H475L substitution that is present in S288c and has been shown to significantly reduce HO activity (Meiron et al. 1995; Ekino et al. 1999). In addition, an in-frame deletion between 8 bp direct repeats within the gene removed 36 amino acid residues from the DNA binding domain (residues 524-559). These mutations likely eliminate the endonuclease activity of HO, explaining the heterothallic life cycle of JAY270.

Flocculation is a mechanism through which yeast cells aggregate to form clumps, or flocs. While flocculation is a desirable trait in brewing, it is problematic in sugar cane batch fermentation because it significantly slows fermentation kinetics, and can cause excessive foaming and clogging of pipes. One of the attractive properties of JAY270 is that it rarely flocculates and produces very little foam in fermentation tanks (Basso et al. 2008). The JAY291 haploid is also non-flocculant. We examined the sequences of genes encoding cell surface adhesins involved in flocculation (*FLO1*, *FLO5*, *FLO11*, and others) and found that they are present in JAY291, although, due to their internal repeat structures (Verstrepen et al. 2004), these sequences were sometimes split between two contigs in our assembly. In addition, we looked at the *FLO8* gene, a positive regulator of flocculation. In S288c, *FLO8* contains a nonsense mutation (*flo8-1*) that renders the cells non-flocculant (Liu et al. 1996). The *FLO8* gene in JAY291 appears to be functional, indicating that the non-flocculant phenotype of JAY291 involves a block in a different step of the pathway. Interestingly, among 120 F₂ progeny from the S288c x JAY291 cross described above (Fig. 2) we observed 19 individuals that flocculated in fermentation media and 101 that did not; roughly a 1:7 phenotypic ratio expected for a trait controlled by three unlinked genes ($\chi^2 = 1.22$; $p = 0.27$). We genotyped the flocculant progeny by PCR and restriction digest and observed that all 19 flocculant progeny inherited the *FLO8* allele from JAY291, therefore confirming the requirement of a functional Flo8p for flocculation.

In the S288c genome, there is a Ty1 retrotransposon insertion in the *HAP1* gene (Gaisne et al. 1999). *HAP1* encodes a heme-mediated transcription factor that controls the expression of genes involved in fermentation (Mense and Zhang 2006), and in the biosynthesis of ergosterol which is important for ethanol tolerance (Inoue et al. 2000). As in other industrial strains (Tamura et al. 2004), this interrupting Ty1 element is absent in JAY291. Consistent with this

result, we found that several of the Hap1p regulated genes were expressed at a higher level in JAY270 than in S288c (Supplemental Table 1).

Another gene important for industrial yeast strains is *MIP1*, a nuclear-encoded mitochondrial DNA polymerase (Foury 1989). A polymorphism present in Mip1p (Mip1p-661A) has been shown recently to be responsible for the high frequency of spontaneous cytoplasmic respiratory mutations (ρ^- *petites*) observed in S288c (Baruffini et al. 2007). The *MIP1* sequence in JAY291 contains five non-synonymous substitutions relative to S288c, including the Mip1p-661T variant, which is associated with a low frequency of *petites*. This desirable trait was confirmed in semi-quantitative *petite* mutation assays in which JAY270 produced *petite* colonies at a frequency about five fold lower than S288c (data not shown).

Efficient sucrose utilization is essential in sugar cane fermentation. *S. cerevisiae* strains have long been known to vary in the number of copies of the *SUC* genes coding for invertase, the enzyme that breaks down sucrose in glucose and fructose which are then fermented to produce ethanol (Carlson and Botstein 1983). S288c only has one copy, *SUC2*, the only non-telomeric member of this gene family. Since JAY270 thrives in the sugar cane extract, we expected to find several *SUC* copies in its genome. Instead, the only invertase gene found in the genome sequence of JAY291 was *SUC2*. Another gene we did not find in JAY291 was *RTM1* (Ness and Aigle 1995), involved in resistance to toxins found in molasses, a byproduct of sugar production sometimes used in bioethanol fermentation. Incidentally, *RTM1* is always found clustered with the telomeric copies of *SUC* genes (Ness and Aigle 1995).

A comprehensive list of all genes found in both JAY291 and S288c is presented in Supplemental Table 1, including the local density of SNPs inside the coding sequences and in the upstream regulatory regions, and the relative level of gene expression between these strains measured with cDNA microarrays. In addition to the shared genes discussed above, it is possible that the genes present in JAY291 but absent in S288c also contribute to the properties of JAY270. At least sixteen such genes have been identified in the genome of JAY291 (Supplemental Table 1). An attractive example of such genes are two copies of putative S-adenosylmethionine-dependent methyltransferases found in the JAY291 genome (contig404-gene1 and contig386-gene1; their predicted protein sequences share 92% similarity). These new genes participate in the same processes as the amplified *SAM3* and *SAM4* genes discussed above. Interestingly, a protein identical to the one coded by contig404-gene1 has been recently identified and found to be specific to the West African lineage of *S. cerevisiae* strains (hypothetical protein 5; Liti et al. 2009).

DISCUSSION

Diverse genome structure of JAY270 and its implications for genome evolution

Our molecular analysis of the JAY270 diploid revealed that its genome is highly heterogeneous, both structurally and at the nucleotide level. Thus far, the genomic analysis of *S. cerevisiae* had been mostly limited to haploids, often derived from homothallic diploid strains. In this type of life cycle, immediately after meiosis, the haploid spores undergo mating type switching, followed by self mating to restore a diploid state (Herskowitz 1988). This cycle generates diploids that are homozygous for the entire genome, except at the *MAT* locus. In contrast, heterothallic spores, such as those derived from JAY270, are unable to switch mating type because of a mutation in the *HO* gene. The spores restore the diploid state by mating to spores of the opposite mating type, either siblings from the same ascus or by outcrossing (Knop 2006). The heterozygous JAY270 genome richly illustrates how this process is effective in preserving and amplifying genetic diversity. It is important to note, however, that both the homothallic and heterothallic life cycles require the production of viable haploid spores in meiosis, and thus both types of strains are expected to continuously remove from the gene pool recessive lethal mutations and chromosomal rearrangements associated with the loss of essential genes.

We have found that the genomes of JAY270 and S288c are collinear through most regions and hybrids generated by crosses between these genetic backgrounds have good spore viability. Structural variation, though abundant, was limited to the end sections of chromosomes. A similar pattern has been previously observed in microarray-based surveys of genome variation among different *S. cerevisiae* strains (Winzeler et al. 2003; Carreto et al. 2008; Schacherer et al. 2009), which found that most polymorphisms, including SNPs and CNVs, were more common within 25 kb of the telomeres. In these studies, however, the full extent of the structural variation in these regions could not be revealed because the microarrays used were based on the S288c genome, and regions of high sequence divergence or sequences that were absent in the reference strain could not be examined. This gap has been recently filled by a whole genome sequencing-based survey of dozens of *S. cerevisiae* and *S. paradoxus* strains (Liti et al. 2009). This study identified DNA sequences, including 38 genes, that are present in some of these strains but are absent in the reference S288c genome. Importantly, these new genes are mostly subtelomeric.

The peripheral regions of chromosomes in *S. cerevisiae* typically contain genes that participate in alternative carbon source and vitamin metabolisms, ion and amino acid transport,

flocculation and other processes that are not essential for viability. In the S288c genome, ~8% of all genes are located at sites distal to essential genes (Supplemental Table 1). Some of these genes, like *SNO/SNZ* and *SAM*, are functionally clustered. Previous studies have shown that there are very few functionally clustered genes in *S. cerevisiae*, and that those which are clustered tend to be associated with specific functions that differ between strains (Hall and Dietrich 2007). Analogous distributions of non-essential genes are also found in other species. For example, in *Plasmodium falciparum*, the peripheral regions encode variable surface antigen genes (contingency genes) that help this parasite evade the host's immune defenses (Freitas-Junior et al. 2000; Barry et al. 2003). Many of the genes found in such regions are members of multi-copy gene families that allow structural variation to occur through ectopic homologous recombination events similar to those detected in JAY270. In addition to their repetitive nature, the telomere-proximal regions of chromosomes are often spatially co-localized at the nuclear envelope, forming clusters of chromatin that are believed to facilitate ectopic recombination interactions (Palladino et al. 1993; Gasser et al. 2004). Pryde *et al.* (1997) integrated these observations to formulate a model in which the yeast genome is structured in chromosomal sectors that are either rigid (central core) or plastic (peripheries). A schematic model of this concept as it relates to a set of *S. cerevisiae* haplotypes is shown in Fig. 8.

We recently demonstrated a mechanism in diploid strains that explains how repetitive elements participate in ectopic homologous recombination through their differential interactions with DNA double strand breaks (DSBs; Argueso et al. 2008). While lesions in single-copy regions of the genome are efficiently repaired to restore the original structures, DSBs that occur within repetitive sequences often lead to chromosomal rearrangements by engaging in homologous recombination with ectopic repeats. This experiment imposed no restrictions regarding the fitness effect or the haploid viability associated with the rearrangements, thus, we identified re-structuring events spanning the entire *S. cerevisiae* genome. Because most rearrangements were non-reciprocal and involved regions containing essential genes, it is likely that those events would be deleterious in meiosis, generating inviable spores and therefore eliminating them from the population. In contrast, the naturally-occurring rearrangements observed in JAY270 are not detrimental, since the translocations do not span genes essential for viability. At least six chromosomes in JAY270 were polymorphic in length and yet the diploid had excellent spore viability. Further, crosses between spores derived from JAY270 and S288c generated hybrid diploids in which structural differences were present in virtually every homolog pair. Despite their heterogeneous karyotype, these hybrids produced mostly viable spores (75-

95%). We propose that all DSBs that occur in repetitive elements of the genome are able to promote ectopic recombination and generate chromosomal rearrangements. Those that span essential genes result in poor spore viability and are, therefore, eventually eliminated from the population, whereas those that involve the peripheral chromosome regions do not affect spore viability and may lead to increased fitness.

A distinguishing feature of the JAY270 genome architecture that may contribute to its fitness in competitive environments is the hemizygous distribution of useful stress tolerance genes. It is possible that this configuration offers alternative mechanisms for diploid cells to adjust the copy number of these genes through mitotic crossover, break-induced replication, or meiotic segregation followed by re-mating of sibling haploid spores (Paques and Haber 1999; Knop 2006; Llorente et al. 2008). For example, these mechanisms could, in just a few generations, produce diploids containing 0 to 4 copies of *MPR* genes and 6 to 12 copies of *SNO/SNZ* genes. Notably, once the homozygous state is achieved, it may be reversed over time through outcrossing or ectopic homologous recombination between neighboring repeats. This scenario may explain how the highly plastic genome structure at the peripheral regions of chromosomes could enable *S. cerevisiae* to rapidly respond to a changing environment.

Finally, it is interesting to contrast the pattern of spatially restricted chromosomal rearrangements in *S. cerevisiae* to the much more radical chromosomal rearrangements recently described in the haploid asexual yeast pathogen *Candida glabrata* (Polakova et al. 2009). These very complex genome reconfiguration events included chromosome fusions, large translocations, aneuploidy and even possible circularizations, that are completely incompatible with meiosis. These authors proposed that *C. glabrata* has taken genome re-structuring to the extreme, sacrificing its sexual life cycle to fully explore the adaptation potential provided by a highly unstable genome structure. We believe that *S. cerevisiae* has found the middle ground by confining the genome rearrangements to the ends of chromosomes.

Implications of highly adapted industrial strains to future bioenergy technologies.

In addition to the insights into genome evolution discussed above, the analysis of the JAY270 genome offers a number of opportunities for the development of a new generation of industrial yeast strains. The finding that JAY270 is genetically compatible with laboratory strains indicates that the full arsenal of research tools developed to study yeast as a model system can be readily applied to the modification of this naturally adapted strain. We envision the use of JAY270 as an ideal delivery vehicle for future renewable energy technologies. One of the most

active fields of investigation in the effort to develop these technologies has been the search for strategies to efficiently break down cellulosic feedstocks and to utilize its complex derived sugars in bioethanol fermentation (Gura 2009). These advances would dramatically improve the overall energy balance in bioethanol production. While the early signs of success in this effort are very encouraging, it is important to remember that even when the solutions for this problem are found, there will be a need to efficiently deliver these technologies to the industrial setting. Decades of experience in the development of industrial strains have convincingly demonstrated that any large-scale attempt to introduce genetically modified yeast strains in the bioethanol industry will be futile, unless they are based on naturally adapted strains. Therefore, strains such as JAY270 are likely to play a key role in facilitating the transition from laboratory technological breakthroughs to industrial scale field applications.

METHODS

Growth media and yeast strains. Yeast culture, mating, sporulation, and tetrad dissection were conducted using standard procedures (Rose et al. 1990). Menadione was obtained from Sigma-Aldrich. The PE-2 stock used to purify JAY270 was from our own laboratory collection, which was derived from commercially available dry active yeast sold to bioethanol distilleries (LNF-Latino Americana). The S288c-isogenic strain used as a control in the fermentation assays in Fig. 1a was the BY4741 haploid (Giaever et al. 2002). For the initial genetic characterization of JAY270, the S288c-isogenic strains used were from the FY series (Winston et al. 1995). FY833 (*Mata*, *ura3-52*, *leu2Δ1*, *trp1Δ63*, *his3Δ200*, *lys2Δ202*) and FY834 (*Mata α* , *ura3-52*, *leu2Δ1*, *trp1Δ63*, *his3Δ200*, *lys2Δ202*) were crossed to the JAY270-derived haploid spores JAY289 (*Mata*), JAY290 (*Mata α*), JAY291 (*Mata*), and JAY292 (*Mata α*) to generate hybrid diploids. FY23 (*Mata*, *ura3-52*, *leu2Δ1*, *trp1Δ63*) and FY86 (*Mata α* , *ura3-52*, *leu2Δ1*, *his3Δ200*) were crossed to each other to generate the JAY309 diploid used as a control in comparative stress tolerance assays (Fig. 1b, and Supplemental Fig. 2 and 3). Finally, the S288c-isogenic strains used in the quantitative genetic analysis described in Fig. 2 were S1 and S97 (Steinmetz et al. 2002). Genomic DNA was also prepared from RM11-1a and YJM789 strains.

Fermentation assays. The fermentation assays shown in Fig. 1a were carried out by inoculating 4 g of fresh yeast cells in 36 ml of autoclaved sugar cane extract (pH4.0, 18.2% Total Reducing Sugars - mostly sucrose). Fermentation cultures were incubated at 30C without agitation for exactly 15 hours. The cells were recovered by centrifugation and the supernatant

was distilled for the measurement of ethanol by densimetry. Cells were recycled for five consecutive fermentation cycles, and four independent replicates were analyzed for each genotype. In each replicate, only the data from cycles 2 to 5 were considered, since the data from the first cycle (adaptation round) is usually highly variable. The final ethanol concentration measurements from cycles 2 to 5 from all four replicates were combined to generate the average and standard error shown.

The fermentation assays used in the quantitative genetic analysis (Fig. 2) were carried out by inoculating 0.08 g of fresh yeast cells in 8 ml of media containing 0.5% yeast extract, 1% peptone, and 10% glucose. Cells were pre-grown in the same media for two cycles to provide an adaptation round. The cell mass concentration was normalized to 1% prior to the third cycle which was used to record cell growth and ethanol production kinetics. Cultures were incubated at 30C in a slow rotating drum to avoid decantation. Samples were taken at regular intervals and centrifuged in pre-weighted tubes to determine the weight of the cell pellets. The supernatant was frozen and later used to determine the ethanol concentration using the BioChain (Hayward, CA) Saccharide Removal and Ethanol Assay kits in sequence according to the manufacturer's recommendations.

Microarray and physical analysis of DNA. CGH-array and Band-array experiments, and PFGE procedures were conducted as previously described (Argueso et al. 2008). Image tracing of PFGE bands was obtained using BioRad Quantity One software. All PCR reactions were conducted using BioRad iProof High-Fidelity DNA polymerase. For the differential gene expression experiments shown in Supplemental Table 1, total RNA was extracted from exponentially growing cells of JAY309 and JAY270. Three independent biological replicates were grown in YP 2% glucose liquid media, with 30C incubation and 200 RPM agitation. First strand cDNA was synthesized and labeled with dUTP-Cy3 or dUTP-Cy5, respectively, using the Invitrogen SuperScript Direct labeling kit. Labeled cDNAs were then competitively hybridized to microarrays containing all S288c ORFs. The Log2 (Cy5/Cy3) signal for all three replicates and average for each microarray probe is shown. The CGH/Band-array, and the differential gene expression microarray data are available under GEO accession numbers GSE14601 and GSE17578, respectively.

Parallel sequencing and genome assembly. Total genomic DNA from JAY291 was isolated using Qiagen Maxi kit. We performed sequencing in the Illumina Genome Analyzer II and

Roche Genome Sequencer FLX (454) platforms (Table 1), using standard manufacturer recommended sample preparation procedures. The Illumina and 454 reads were assembled into longer contigs using *de novo* assembler Edena (Hernandez et al. 2008) and Newbler (Roche), respectively. The hybrid assembly was performed through the combination of 3581 Illumina contigs and 1259 454 contigs resulting in 452 hybrid contigs (pipeline developed in our group; Zorro: A hybrid assembly pipeline for next-gen sequencing data; G. G. L. Costa, R. O. V., M. F. C., J. M. C. Mondego, M. J. Guiltinam, S. C. Schuster, J. E. Carlson, P. A. M., J. L. A., L. W. Meinhardt and G. A. G. P.; *manuscript in preparation*). Briefly, the zorro pipeline consists of (1) masking repeat regions in the contigs, (2) overlap detection, (3) unmasking repeat regions and (4) assembly of hybrid contigs. In phase (1), the repeat regions were determined based on counting occurrences of k-mers in the Roche 454 reads, the assembler then masks k-mers in the contigs that occur at high frequency. The absence of repeats produces correct overlap detection in phase (2). The contigs are then unmasked in phase (3), and the correct hybrid assembly can be obtained in (4), by merging all overlapping contigs into hybrid contigs. The overlaps detection and consensus generation were performed using the minimus package (Sommer et al. 2007). The hybrid contigs were ordered and oriented with Bambus program (Pop et al. 2004) using paired-end information and manual verification producing 427 scaffolds. These sequences were aligned to the reference genome (S288c) using Mummer package (Kurtz et al. 2004). The genome comparisons were visualized using Artemis Comparison Tool - ACT (Carver et al. 2005). *ab initio* and extrinsic gene predictions were obtained using Augustus (Stanke and Morgenstern 2005) and Exonerate (Slater and Birney 2005), respectively. For the Augustus analysis, we used the training sets available for S288c, and for Exonerate we used the reference curated ORFs (SGD; <http://www.yeastgenome.org>) that were aligned into our hybrid assembly. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ACFL000000000. The version described in this paper is the first version, ACFL01000000.

ACKNOWLEDGEMENTS

We are grateful to Chris Vinci, Steven Clarke, Dina Raveh, Joseph Heitman, and Eric Alani for useful discussions and comments on the manuscript. We thank CENAPAD-SP (Centro Nacional de Processamento de Alto Desempenho de São Paulo) for data processing support. Research in the Pereira, Tavares, Alcarde and Andrietta laboratories is supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, FAPESP (Fundação de

Amparo à Pesquisa do Estado de São Paulo), and CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior). Research in the Pereira laboratory is also supported by BRASKEM and ETH Bioenergia. Work in the Petes laboratory is funded by NIH grants GM52319 and GM24110.

REFERENCES

- Argueso, J.L., Westmoreland, J., Mieczkowski, P.A., Gawel, M., Petes, T.D., and Resnick, M.A. 2008. Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc.Natl.Acad.Sci.U.S.A* **105**:11845-11850
- Barry, J.D., Ginger, M.L., Burton, P., and McCulloch, R. 2003. Why are parasite contingency genes often associated with telomeres? *Int.J.Parasitol.* **33**:29-45
- Baruffini, E., Lodi, T., Dallabona, C., and Foury, F. 2007. A single nucleotide polymorphism in the DNA polymerase gamma gene of *Saccharomyces cerevisiae* laboratory strains is responsible for increased mitochondrial DNA mutability. *Genetics* **177**:1227-1231
- Basso, L.C., de Amorim, H.V., de Oliveira, A.J., and Lopes, M.L. 2008. Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Res.*
- Borneman, A.R., Forgan, A.H., Pretorius, I.S., and Chambers, P.J. 2008. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Res.* **8**:1185-1195
- Carlson, M. and Botstein, D. 1983. Organization of the *SUC* gene family in *Saccharomyces*. *Mol.Cell Biol.* **3**:351-359
- Carreto, L., Eiriz, M.F., Gomes, A.C., Pereira, P.M., Schuller, D., and Santos, M.A. 2008. Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC Genomics* **9**:524
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics.* **21**:3422-3423
- da Silva, E.A., dos Santos, S.K.B., Resende, A.D., de Moraes, J.O.F., de Moraes, M.A., and Simoes, D.A. 2005. Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **88**:13-23
- Doniger, S.W., Kim, H.S., Swain, D., Corcuera, D., Williams, M., Yang, S.P., and Fay, J.C. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS.Genet.* **4**:e1000183
- Du, X. and Takagi, H. 2007. N-Acetyltransferase Mpr1 confers ethanol tolerance on *Saccharomyces cerevisiae* by reducing reactive oxygen species. *Appl.Microbiol.Biotechnol.* **75**:1343-1351
- Ehrenshaft, M., Bilski, P., Li, M.Y., Chignell, C.F., and Daub, M.E. 1999. A highly conserved

- sequence is a novel gene involved in de novo vitamin B6 biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **96**:9374-9378
- Ekino, K., Kwon, I., Goto, M., Yoshino, S., and Furukawa, K. 1999. Functional analysis of *HO* gene in delayed homothallism in *Saccharomyces cerevisiae* wy2. *Yeast* **15**:451-458
- Faddah, D.A., Ganko, E.W., McCoach, C., Pickrell, J.K., Hanlon, S.E., Mann, F.G., Mieczkowska, J.O., Jones, C.D., Lieb, J.D., and Vision, T.J. 2009. Systematic identification of balanced transposition polymorphisms in *Saccharomyces cerevisiae*. *PLoS.Genet.* **5**:e1000502
- Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., and Louis, E.J. 2000. Chromosomal evolution in *Saccharomyces*. *Nature* **405**:451-454
- Foury, F. 1989. Cloning and sequencing of the nuclear gene *MIP1* encoding the catalytic subunit of the yeast mitochondrial DNA polymerase. *J.Biol.Chem.* **264**:20552-20560
- Fox, T.D., Folley, L.S., Mulero, J.J., McMullin, T.W., Thorsness, P.E., Hedin, L.O., and Costanzo, M.C. 1991. Analysis and manipulation of yeast mitochondrial genes. *Methods in Enzymology* **194**:149-165
- Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E., and Scherf, A. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**:1018-1022
- Gaisne, M., Becam, A.M., Verdiere, J., and Herbert, C.J. 1999. A 'natural' mutation in *Saccharomyces cerevisiae* strains derived from S288c affects the complex regulatory gene *HAP1* (*CYP1*). *Curr.Genet.* **36**:195-200
- Gasser, S.M., Hediger, F., Taddei, A., Neumann, F.R., and Gartenberg, M.R. 2004. The function of telomere clustering in yeast: the circe effect. *Cold Spring Harb.Symp.Quant.Biol.* **69**:327-337
- Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**:387-391
- Goffeau, A. 2004. Evolutionary genomics: seeing double. *Nature* **430**:25-26
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. Life with 6000 genes. *Science* **274**:546, 563-546, 567
- Goldemberg, J. 2007. Ethanol for a sustainable energy future. *Science* **315**:808-810
- Gordon, J.L., Byrne, K.P., and Wolfe, K.H. 2009. Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS.Genet.* **5**:e1000485
- Gura, T. 2009. Driving Biofuels from Field to Fuel Tank. *Cell* **138**:9-12

- Hall, C. and Dietrich, F.S. 2007. The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics* **177**:2293-2307
- Hernandez, D., Francois, P., Farinelli, L., Osteras, M., and Schrenzel, J. 2008. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res.* **18**:802-809
- Herskowitz, I. 1988. Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol.Rev.* **52**:536-553
- Inoue, T., Iefuji, H., Fujii, T., Soga, H., and Satoh, K. 2000. Cloning and characterization of a gene complementing the mutation of an ethanol-sensitive mutant of sake yeast. *Bioscience Biotechnology and Biochemistry* **64**:229-236
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**:241-254
- Knop, M. 2006. Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. *Bioessays* **28**:696-708
- Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* **5**:R12
- Landolfo, S., Politi, H., Angeozzi, D., and Mannazzu, I. 2008. ROS accumulation and oxidative damage to cell structures in *Saccharomyces cerevisiae* wine strains during fermentation of high-sugar-containing medium. *Biochimica et Biophysica Acta-General Subjects* **1780**:892-898
- Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., et al. 2009. Population genomics of domestic and wild yeasts. *Nature* **458**:337-341
- Liu, H.P., Styles, C.A., and Fink, G.R. 1996. *Saccharomyces cerevisiae* S288C has a mutation in *FL08* a gene required for filamentous growth. *Genetics* **144**:967-978
- Llorente, B., Smith, C.E., and Symington, L.S. 2008. Break-induced replication: what is it and what is it for? *Cell Cycle* **7**:859-864
- Meiron, H., Nahon, E., and Raveh, D. 1995. Identification of the heterothallic mutation in HO-endonuclease of *S. cerevisiae* using HO/ho chimeric genes. *Curr.Genet.* **28**:367-373
- Mense, S.M. and Zhang, L. 2006. Heme: a versatile signaling molecule controlling the activities of diverse regulators ranging from transcription factors to MAP kinases. *Cell Res.* **16**:681-692
- Ness, F. and Aigle, M. 1995. *RTM1*: a member of a new family of telomeric repeated genes in yeast. *Genetics* **140**:945-956

- Padilla, P.A., Fuge, E.K., Crawford, M.E., Errett, A., and Werner-Washburne, M. 1998. The highly conserved, coregulated *SNO* and *SNZ* gene families in *Saccharomyces cerevisiae* respond to nutrient limitation. *Journal of Bacteriology* **180**:5718-5726
- Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., and Gasser, S.M. 1993. SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell* **75**:543-555
- Paques, F. and Haber, J.E. 1999. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol.Mol.Biol.Rev.* **63**:349-404
- Polakova, S., Blume, C., Zarate, J.A., Mentel, M., Jorck-Ramberg, D., Stenderup, J., and Piskur, J. 2009. Formation of new chromosomes as a virulence mechanism in yeast *Candida glabrata*. *Proc.Natl.Acad.Sci.U.S.A* **106**:2688-2693
- Pop, M., Kosack, D.S., and Salzberg, S.L. 2004. Hierarchical scaffolding with Bambus. *Genome Res.* **14**:149-159
- Pryde, F.E., Gorham, H.C., and Louis, E.J. 1997. Chromosome ends: all the same under their caps. *Curr.Opin.Genet.Dev.* **7**:822-828
- Robertson, G.P., Dale, V.H., Doering, O.C., Hamburg, S.P., Melillo, J.M., Wander, M.M., Parton, W.J., Adler, P.R., Barney, J.N., Cruse, R.M., et al. 2008. Agriculture. Sustainable biofuels redux. *Science* **322**:49-50
- Rodriguez-Navarro, S., Llorente, B., Rodriguez-Manzanegue, M.T., Ramne, A., Uber, G., Marchesan, D., Dujon, B., Herrero, E., Sunnerhagen, P., and Perez-Ortin, J. 2002. Functional analysis of yeast gene families involved in metabolism of vitamins B-1 and B-6. *Yeast* **19**:1261-1276
- Rose, M.D., Winston, F., and Hieter, P. 1990. *Methods in yeast genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanderson, K. 2006. US biofuels: a field in ferment. *Nature* **444**:673-676
- Scannell, D.R., Butler, G., and Wolfe, K.H. 2007. Yeast genome evolution--the origin of the species. *Yeast* **24**:929-942
- Schacherer, J., Shapiro, J.A., Ruderfer, D.M., and Kruglyak, L. 2009. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**:342-345
- Sinha, H., David, L., Pascon, R.C., Clauser-Munster, S., Krishnakumar, S., Nguyen, M., Shi, G., Dean, J., Davis, R.W., Oefner, P.J., et al. 2008. Sequential elimination of major-effect contributors identifies additional quantitative trait Loci conditioning high-temperature growth in yeast. *Genetics* **180**:1661-1670
- Sinha, H., Nicholson, B.P., Steinmetz, L.M., and McCusker, J.H. 2006. Complex genetic interactions in a quantitative trait locus. *PLoS.Genet.* **2**:e13

- Slater, G.S. and Birney, E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. **6**:31
- Sommer, D.D., Delcher, A.L., Salzberg, S.L., and Pop, M. 2007. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics*. **8**:64
- Stanke, M. and Morgenstern, B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**:W465-W467
- Steinmetz, L.M., Sinha, H., Richards, D.R., Spiegelman, J.I., Oefner, P.J., McCusker, J.H., and Davis, R.W. 2002. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**:326-330
- Takagi, H., Shichiri, M., Takemura, M., Mohri, M., and Nakamori, S. 2000. *Saccharomyces cerevisiae* sigma 1278b has novel genes of the N-acetyltransferase gene superfamily required for L-proline analogue resistance. *J.Bacteriol.* **182**:4249-4256
- Tamura, K., Gu, Y., Wang, Q., Yamada, T., Ito, K., and Shimoi, H. 2004. A *hap1* mutation in a laboratory strain of *Saccharomyces cerevisiae* results in decreased expression of ergosterol-related genes and cellular ergosterol content compared to sake yeast. *J.Biosci.Bioeng.* **98**:159-166
- Tavares, F.C.A. and Echeverrigaray, S. 1987. Yeast Breeding for fuel production. In *Biochemistry and Biology of Industrial Yeasts* (eds. G. Stewart, I. Russell, R. Klein, R. Hiebish), pp. 59-80. CRC, Florida.
- Verstrepen, K.J., Reynolds, T.B., and Fink, G.R. 2004. Origins of variation in the fungal cell surface. *Nat.Rev.Microbiol.* **2**:533-540
- Vinci, C.R. and Clarke, S.G. 2007. Recognition of age-damaged (R,S)-adenosyl-L-methionine by two methyltransferases in the yeast *Saccharomyces cerevisiae*. *J.Biol.Chem.* **282**:8604-8612
- Wang, P., Kim, Y., Pollack, J., Narasimhan, B., and Tibshirani, R. 2005. A method for calling gains and losses in array CGH data. *Biostatistics*. **6**:45-58
- Wei, W., McCusker, J.H., Hyman, R.W., Jones, T., Ning, Y., Cao, Z., Gu, Z., Bruno, D., Miranda, M., Nguyen, M., et al. 2007. Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc.Natl.Acad.Sci.U.S.A* **104**:12825-12830
- Wheals, A.E., Basso, L.C., Alves, D.M., and Amorim, H.V. 1999. Fuel ethanol after 25 years. *Trends Biotechnol.* **17**:482-487
- Winston, F., Dollard, C., and RicuperoHovasse, S.L. 1995. Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* **11**:53-55
- Winzeler, E.A., Castillo-Davis, C.I., Oshiro, G., Liang, D., Richards, D.R., Zhou, Y., and Hartl, D.L. 2003. Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics* **163**:79-89

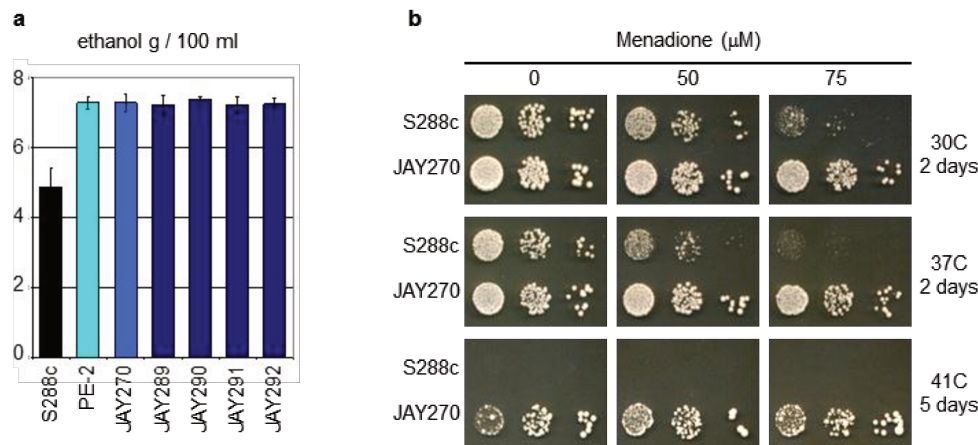


Figure 1. Comparative phenotypic analysis of JAY270 and S288c.

a. Sugar cane extract fermentation assays (~18% sucrose). The ethanol concentration shown is the average reached in four fermentation assay repetitions, each comprised of four consecutive 15 hours of fermentation at 30C with cell recycling (see Materials and Methods for details). **b.** Reactive oxygen (menadione) and temperature effects on colony growth. Ten-fold serial dilutions of saturated cultures spotted (5 μ l) in rich medium (YPD) containing various concentrations of menadione were incubated at the indicated temperatures, and scanned after the indicated incubation period. The S288c-isogenic control strain used in **(a)** was the BY4741 haploid; and in **(b)** was the JAY309 diploid.

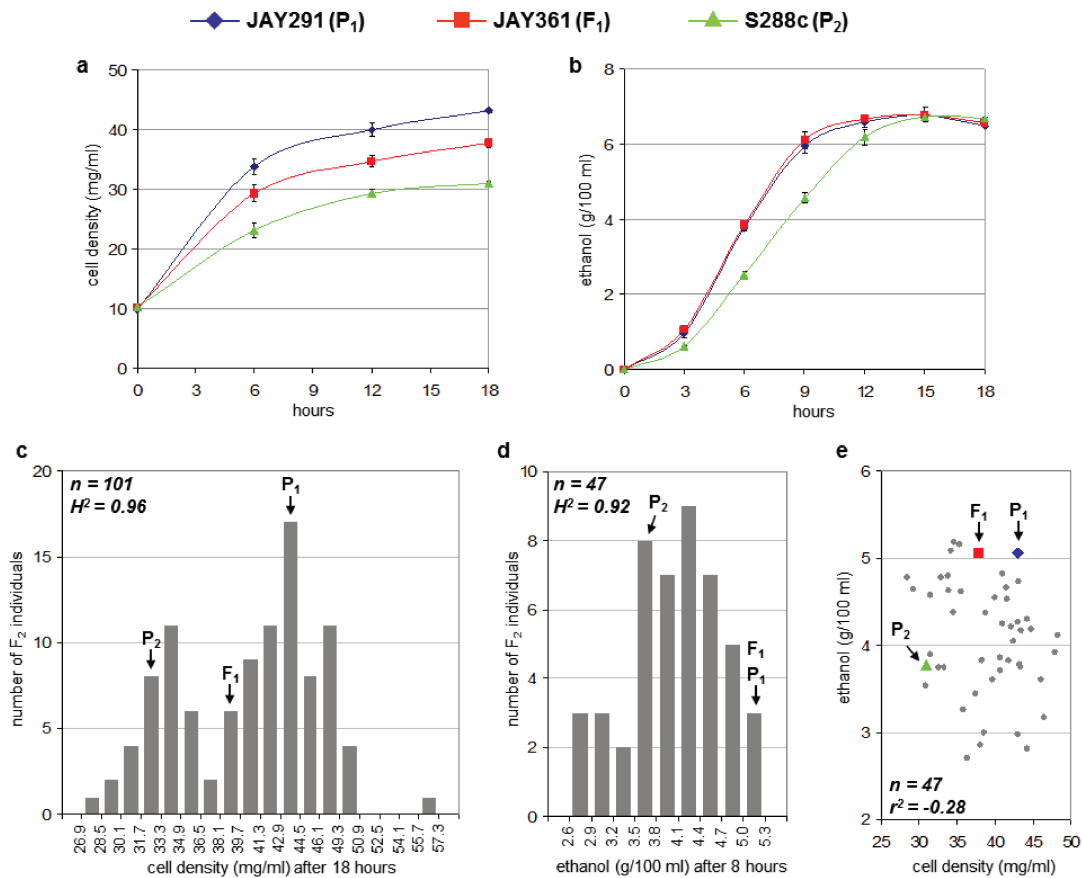


Figure 2. Genetic analysis of cell mass accumulation and ethanol production kinetics.

a and **b**. Kinetics of cell mass accumulation and ethanol production, respectively, during fermentation of rich media with 10% glucose. The results show the average values and standard error for three biological replicates from each strain. The S288c-isogenic strains used in these assays were the *MAT α* haploids S1 (ρ^+) and S97 (ρ^0). The JAY361 diploid was obtained by mating JAY291 and S97; this diploid inherited 100% of its mitochondrial genome from JAY291. **c.** and **d.** show the distributions of cell mass accumulation after 18 hours and ethanol production after 8 hours (respectively) among haploid F₂ spores derived from JAY361. F₂ individuals were grouped in bins according to their phenotypes, and the bars represent the number of individuals in each phenotypic bin. **e.** Scatter plot of cell density (x axis) vs. ethanol concentration (y axis) for all F₂ individuals tested in both assays (gray dots). P₁, F₁ and P₂ indicate the relative phenotypes of JAY291, JAY361 and S1, respectively. *n* is the number of F₂ individuals in each dataset, *H*² is the broad-sense heritability calculated from phenotypic and environmental variances, and *r*² is coefficient of correlation between the two traits analyzed. Only data from non-flocculant F₂ individuals was used in this analysis.

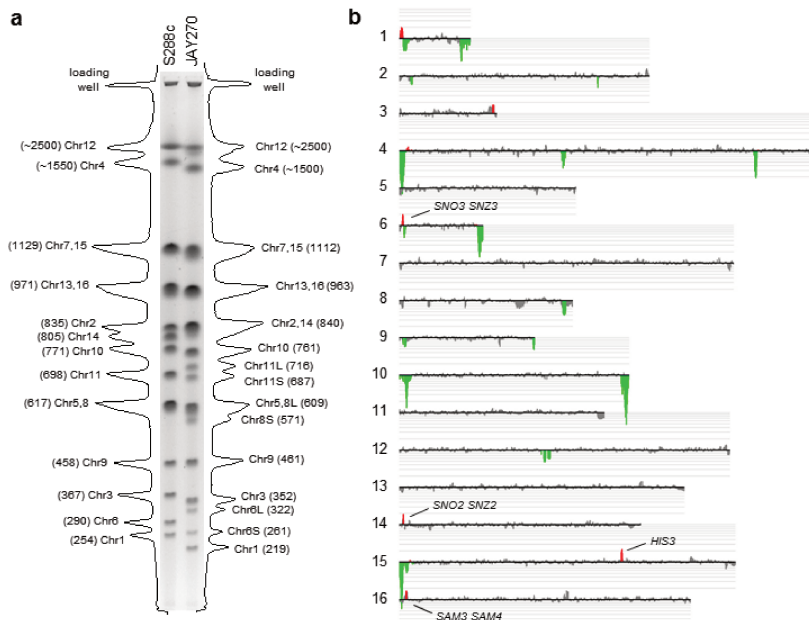


Figure 3. Molecular karyotype and gene copy number variation in JAY270.

a. PFGE and densitometric analysis of individual chromosomal bands in S288c and JAY270. The size of the peaks reflects the intensity of the ethidium bromide staining for each chromosomal band as determined by image analysis using BioRad QuantityOne software. The predicted chromosome sizes (in kb) are shown next to the corresponding chromosomal peaks were determined by comparison to the BioRad I molecular weight ladder (not shown). Note the presence of different sized homologs for Chr6 and Chr11 that appear at lower relative intensities. Contrast the abundance of each Chr6 homolog to the intensity of Chr3 for which both homologs are about the same size, and compare the abundance of each Chr11 homolog to the intensity of Chr10. **b.** CGH-array relative gene dosage plots. Each horizontal line corresponds to a specific S288c chromosome; the signal of each array probe was smoothed in CGH-miner software in a seven-probe sliding window to reduce noise (Wang et al. 2005). Gray areas indicate regions of similar genomic dosage, while positive/red and negative/green peaks indicate genomic regions over- or under-represented, respectively. The amplification signal on the left end of Chr1 included the *SEO1* gene encoding a putative amino-acid permease; the amplification peak on the right end of Chr3 did not include any known genes; all other amplifications are discussed in the text. None of the deletion/under-representation peaks spanned regions containing genes known to be essential in S288c. The S288c-isogenic control strain used in (a) and (b) was the JAY309 diploid.

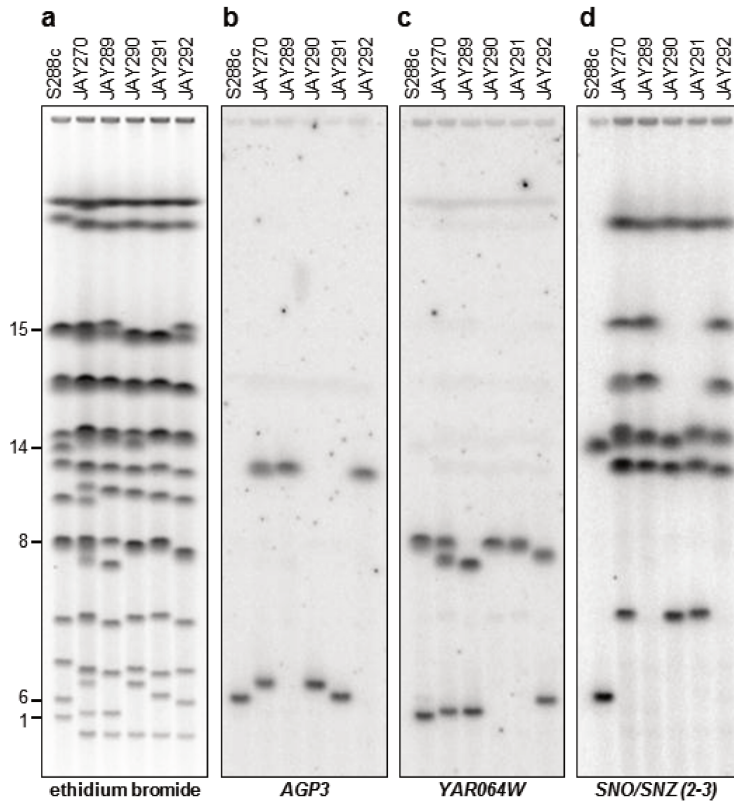


Figure 4. Molecular karyotype of meiotic products and segregation of chromosomal rearrangements.

a. Ethidium bromide staining of a PFGE including the four meiotic spore clones from a JAY270 tetrad. The S288c-isogenic control strain used was the JAY309 diploid. **b**, **c**, and **d** show Southern hybridizations of the PFGE in (**a**), using as probes, respectively, the *AGP3*, *YAR064W*, and *SNO/SNZ* sequences that were PCR-amplified from S288c genomic DNA. The *SNO/SNZ* (2-3) probe detects only the duplicated *SNO2*, *SNZ2*, *SNO3*, and *SNZ3* genes. The diverged single copy genes *SNO1* and *SNZ1* on Chr13 are not detected. The numbers to the left indicate the S288c chromosomes to which these probes hybridized. The *YAR064W* gene is duplicated in S288c and JAY270 at the right end of Chr8.

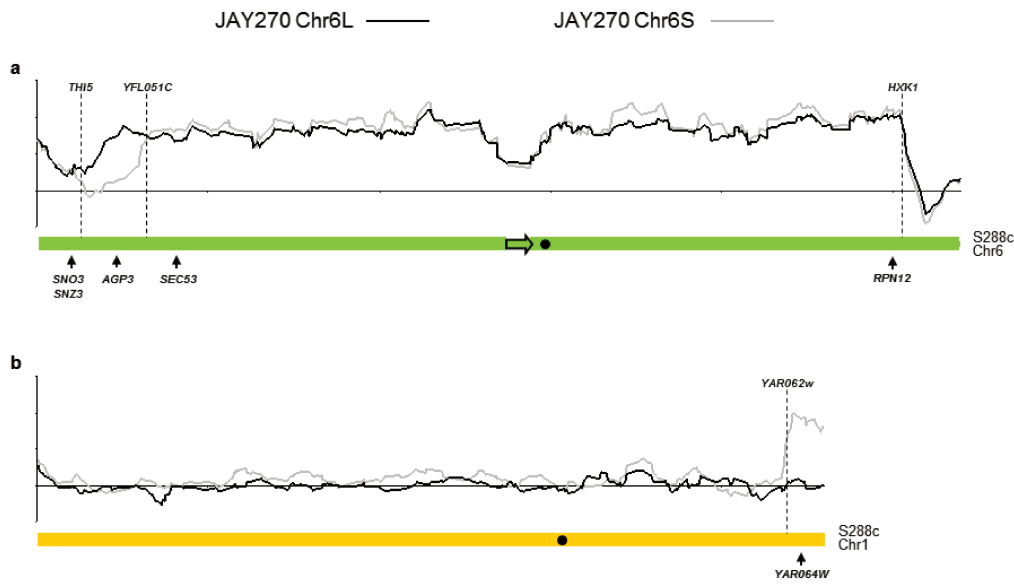


Figure 5. Band-array analysis of Chr6L and Chr6S homologs in JAY270.

The curves indicate the normalized hybridization signal of specific chromosomal DNA samples (y axis) to probes arranged according to their chromosomal coordinates (x axis) in S288c Chr6 (a) and Chr1 (b). The dashed vertical lines indicate the breakpoints where JAY270 Chr6S and Chr6L differ from each other and from the S288c chromosomes. Only the data for Chr6 and Chr1 probes are shown. No hybridization signal was detected for other S288c chromosomes, with the exception of the right end of Chr8 that is essentially identical to Chr1. *SEC53* and *RPN12* indicate the most distal essential genes in S288c Chr6, and *SNO3*, *SNZ3*, *AGP3* and *YAR064W* indicate the position of Southern blot probes used in Fig. 4. The signal valley in the central position of Chr6 corresponds to the position of the *YFLWTy2-1* retrotransposable element (arrow) which is not present in the JAY270 Chr6 homologs. The black circles indicate the centromere positions.

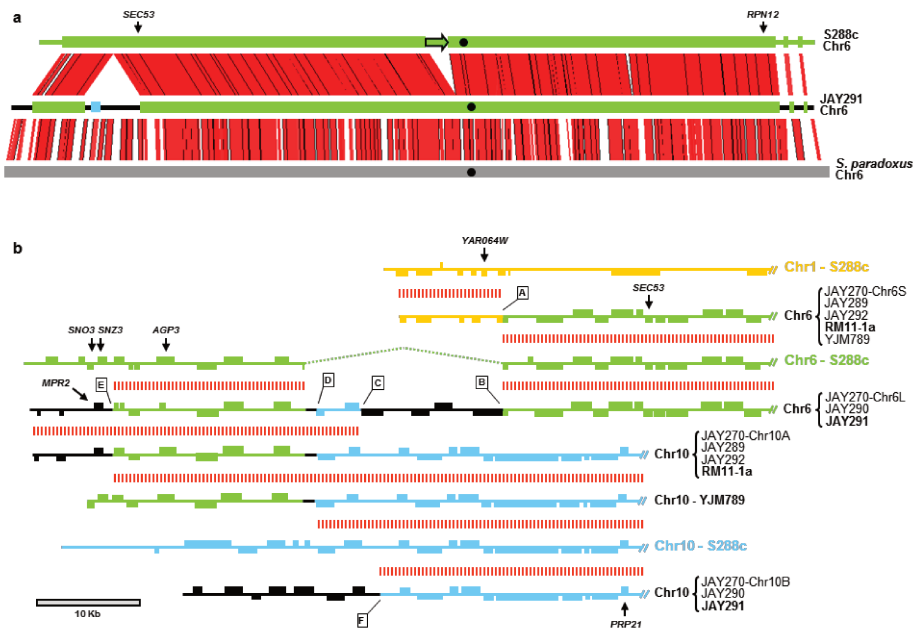


Figure 6. Genome rearrangements near the ends of chromosomes.

a. Full-length Chr6 sequences aligned with Artemis Comparative Tool software (Carver et al. 2005). Red lines connect regions of sequence similarity higher than 85%, gaps in white indicate lower or absent similarity. Green indicates S288c Chr6 sequences and thick areas indicate regions conserved in JAY291, whereas thin areas indicate non-conserved regions. The small segment in Blue is a translocated fragment from S288c Chr10. Black indicates *S. cerevisiae* sequences not found in the S288c genome. The gray thick line corresponds to *S. paradoxus* Chr6 assembled from contigs 346, 345, 344, 434 and 433, from left to right, in this order (Kellis et al. 2003). *SEC53* and *RPN12* indicate the positions of the most distal essential genes in Chr6. Black circles indicate the position of the centromere (*CEN6*) and the arrow represents the *YFLWTy2-1* element. Y' and X sub-telomeric sequences are not represented in this alignment.

b. Multi-strain chromosome alignment around the left end of Chr6. The horizontal lines represent the left ends of the designated chromosomes, with the exception of S288c Chr1 (top line) for which the inverted right end is shown. The source strain for each sequence is indicated to the left in bold, other strains with similar chromosome structures are also indicated. The rectangles represent ORFs, and their positions above and below the central line designate the Watson and Crick orientations, respectively. Chromosomal regions are color coded according to their correspondence to S288c Chr1 (orange), Chr6 (green) and Chr10 (blue), while regions in black correspond to sequences not found in the S288c genome. The dotted line in S288c Chr6 represents a discontinuity in the alignment at the site of a 19.3 kb insertion in JAY291. The

rectangles hatched in red between chromosomes indicate a high level of sequence similarity (>85%; analogous to the red lines in panel a). All sequence similarities are indicated in red except for the left end of YJM789 Chr10 which contains the *SNO3* and *SNZ3* genes and is nearly identical to the collinear region in S288c Chr6. Boxed letters (A-F) indicate the specific chromosomal rearrangements discussed in the text, and *SNO3*, *SNZ3*, *AGP3* and *YAR064W* indicate the positions of Southern blot probes used in Fig. 4. The source sequences for this figure were the complete Chr6 sequences from S288c and JAY291, contigs 1.67 (Chr10) and 1.109 (Chr6) from RM11-1a, contigs 100 (Chr10) and 7 (Chr6) from YJM789, and contig 386 (Chr10) from JAY291. The 10 kb scale bar indicates the size scale for panel b.

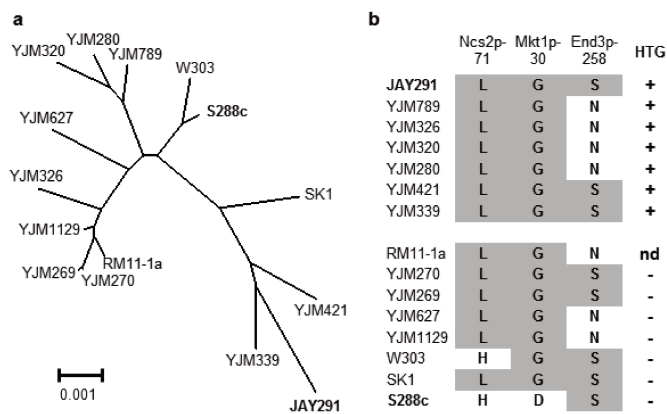


Figure 7. Phylogenetic placement and High-Temperature Growth (HTG) alleles in JAY291.

a. Unrooted phylogenetic tree of 15 *S. cerevisiae* strains based on a 49 kb region from Chr14 containing the three HTG QTLs (Sinha et al. 2008). **b.** HTG QTL allele distribution in various *S. cerevisiae* strains. The three critical amino acid residues implicated in High Temperature Growth are shaded in gray, and HTG (+ , - , nd - not determined) phenotype is shown to the right.

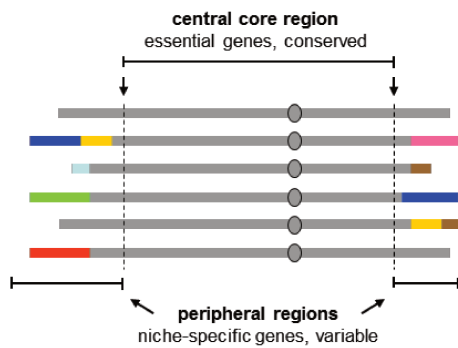


Figure 8. Structural diversity in *S. cerevisiae*: rigid and plastic domains of the genome.

The model depicts a set of homologs of a hypothetical chromosome in several unrelated *S. cerevisiae* strains. The top line (continuous gray) depicts the structural configuration of this chromosome in the first sequenced strain (*i.e.* S288c; reference), whereas the chromosomes shown below represent the rearrangements found in other strains. The diverged structural configurations (colors) in the peripheral regions harbor genes that are not required for viability, but that may contribute to fitness in specific environments. The entire set shares structural conservation in the central core region (delimited by the most distal essential genes - arrows), therefore meiotic crossovers between the unrelated haplotypes can generate new combinations, while remaining compatible with haploid viability.

Table 1. Summary of JAY291 genome sequencing and assembly parameters.

Haplotype	JAY291
Contigs	452
Contig size (average)	25.5 Kb
N50 ^a	65 Kb
Assembly size	11.6 Mb
Coverage depth ^b :	
454 single end	12x
Illumina single end	55x
Illumina paired-ends	95x
Gene models ^c :	
Extrinsic	5864
<i>Ab initio</i>	5488
Total gene models	5880

a. 50% of the genome assembly is represented by contigs of N50 size and longer.

b. Roche/454 reads were about 250 bp, Illumina single end reads were 35 bp, and paired-end reads correspond to two 35 bp reads physically linked by ~125 bp of undetermined sequence.

c. Extrinsic gene models were predicted based on the established S288c ORFs; *ab initio* gene models were predicted without regard to S288c data. 5471 genes were predicted by both models; 392 and 16 genes were found exclusively in the extrinsic and *ab initio* models, respectively.

CAPÍTULO 2

PCR-based molecular markers for the efficient monitoring of yeast cell populations during bioethanol production

Osmar V. Carvalho-Netto ¹; Marcelo F. Carazzolle ¹; Aline Rodrigues ^{1,2}; Welbe O. Bragança ¹; Gustavo G. L. Costa ¹; Juan Lucas Argueso ^{2,*}; and Gonçalo A. G. Pereira ^{1,*}

¹ Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas-SP, Brazil.

² Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins-CO, USA.

* Corresponding authors.

Corresponding authors:

Juan Lucas Argueso, PhD

Department of Environmental and Radiological Health Sciences, Colorado State University, 1618 campus delivery, Fort Collins-CO, 80523, USA.

Phone: 1-970-491-3681; Fax: 1-970-491-0623; email: lucas.argueso@colostate.edu

Gonçalo A. G. Pereira, PhD

Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas-SP, 13083-970, Brazil.

Phone: 55-19-3521-6237; email: goncalo@unicamp.br

KEYWORDS

Saccharomyces cerevisiae; PCR; genotyping; bioethanol.

ABSTRACT

Background: One of the defining features of the fermentation process used in the production of bioethanol from sugarcane feedstock is the dynamic nature of the yeast population. The composition of the microbial population is an important parameter that distilleries must monitor periodically in order to detect the presence of wild contaminants before they can cause productivity losses. Pulsed-field gel electrophoresis (PFGE) is the current standard monitoring method used by the industry. While PFGE is effective, it is also expensive, time-consuming, and technically challenging. The ideal alternative method is PCR-based strain identification, however the marker loci available to date often lack sufficient unique polymorphisms in bioethanol strains.

Results: We compared the coding sequences derived from the genome of the sugarcane bioethanol strain JAY270/PE-2 to those of the reference *Saccharomyces cerevisiae* laboratory strain S288c, and searched for genes containing insertion or deletion polymorphisms sufficiently large to be resolved by conventional agarose gel electrophoresis. We then designed PCR primers flanking eleven of these sites, and used them to amplify differentially sized products between these strains. We analyzed the banding patterns in the most widely adopted sugarcane bioethanol strains and in several indigenous yeast contaminants, and found that our marker set had a high discriminatory power. Complete differentiation of all twenty strains analyzed in this training set was achieved using only four of the markers developed in this study. Subsequently, these four core markers (*YFL024C*, *YKL163W*, *YKL201C*, and *YLL021W*) were used to successfully monitor the yeast cell populations in six sugarcane bioethanol distilleries. We also showed that most of the markers developed here are highly polymorphic among sixty unrelated yeast strains that have their complete genomes sequenced and deposited in public databases.

Conclusion: We have described a set of highly informative PCR-based molecular markers for the monitoring of sugarcane bioethanol yeast cell populations. Compared to PFGE, this method is faster, more affordable, and has higher throughput; and should be easily adoptable by modestly equipped quality control laboratories in bioethanol distilleries. Moreover, our analysis indicated that these markers should also be useful in the identification of yeast strains in other unrelated fermentation processes.

BACKGROUND

In Brazil and other tropical countries that grow sugarcane, ethanol biofuel (bioethanol) is produced through fermentation of sucrose by the yeast *Saccharomyces cerevisiae*, primarily using a fed-batch process (WHEALS *et al.* 1999). This high moisture / high sugar feedstock brings into the industrial plant a substantial microbial load of wild yeasts and bacteria (AMORIM *et al.* 2011). A heat treatment is applied to reduce the number of viable contaminants, but inevitably, some find their way into the fermentation tanks where they have the opportunity to flourish. After the fermentation of each batch of sugarcane juice is completed, the cell population is recycled to ferment the subsequent batches, and this practice is repeated continuously for the duration of the harvest season (8-9 months). The effects of the steady influx of microbial contaminants are exacerbated by the cell recycling regimen widely adopted by the industry to achieve a high rate of feedstock processing: *i.e.*, a dense yeast inoculum (~10% cells by volume) leads to faster fermentations (~10 hours). Consequently, the microbial community in most distilleries is heterogeneous and highly dynamic. The starter strains inoculated in the beginning of the season (typically baker's yeast due to wide market availability and low cost) are rapidly out-competed and replaced by more robust indigenous strains that can endure the stressful conditions encountered during the fermentation (DA SILVA *et al.* 2005).

Over the past fifteen years, indigenous strains that combine excellent fermentative characteristics and high adaptation to the industrial environment have been selected from Brazilian bioethanol distilleries and are used today as starter strains. Among those, the PE-2 and CAT-1 strains are the most widely adopted, and in 2007 were responsible for more than half of the bioethanol produced in Brazil (BASSO *et al.* 2008). However, PE-2 and CAT-1 are exceptions among their indigenous counterparts, which in their majority display undesirable fermentation characteristics, and often result in significant productivity losses if they are allowed to establish a foothold in a distillery (BASSO *et al.* 2008; WHEALS *et al.* 1999). Therefore, effective methods are needed to regularly monitor the microbial population, allowing process managers to promptly identify the presence of invaders and, when necessary, take corrective action.

A wide array of methodologies has been previously described for the molecular identification of *S. cerevisiae* strains including ribosomal DNA or ITS sequencing (KURTZMAN and ROBNETT 1998), RAPD (BOVO *et al.* 2009; GOMES *et al.* 2000), mDNA RFLP (QUEROL *et al.* 1992), PCR-RFLP (SCHULLER *et al.* 2007; SCHULLER *et al.* 2004), real-time PCR (HIERRO *et al.* 2006; MARTORELL *et al.* 2005), microsatellite and inter-delta sequence amplification (LEGRAS and KARST 2003; LEGRAS *et al.* 2005; SCHULLER *et al.* 2007) and karyotyping by pulsed-field gel

electrophoresis (PFGE) (BASSO *et al.* 2008; SCHULLER *et al.* 2007; SCHULLER *et al.* 2004). Despite the variety of available assays, each of them has specific shortcomings that reduce their effectiveness, including laborious sample preparation, poor reproducibility, high complexity of data analysis, and most importantly, lack of polymorphisms in specific industrial strains of interest (GOMES *et al.* 2000; SCHULLER *et al.* 2004).

PCR-based methods are faster than PFGE and have been successfully used to identify *S. cerevisiae* strains using a large number of loci (HENNEQUIN *et al.* 2001; LEGRAS *et al.* 2005; MERCADO *et al.* 2010; PEREZ *et al.* 2001; TECHERA *et al.* 2001). However, the design of such markers has been guided primarily by the identification of low complexity sequence regions (*e.g.* microsatellites) in the reference genome sequence of the S288c laboratory strain (EZOV *et al.* 2006; LEGRAS *et al.* 2005; PEREZ *et al.* 2001), based on the expectation that these regions should be polymorphic in unrelated strains. Unfortunately, this expectation is not always met, and markers designed through this approach often lack sufficient polymorphism between bioethanol strains and common fermentation contaminants. For this reason, the standard monitoring assay used in Brazil has been karyotyping by PFGE. While PFGE has proven useful, this analysis is technically challenging (often outsourced to third party laboratories) and very time consuming (several days may elapse between sample collection and the final read-out). Furthermore, the interpretation of the results is sometimes complicated by the fact that industrial yeast strains can display variable PFGE banding patterns due to chromosomal rearrangements (CARRO and PINA 2001; NADAL *et al.* 1999).

We have previously described the complete genome sequence of a haploid derivative (JAY291) from the JAY270/PE-2 bioethanol strain (ARGUESO *et al.* 2009). We took advantage of this sequence data to design high specificity PCR markers that should be immediately useful to the large number of bioethanol distilleries that use JAY270/PE-2. We carried out a genome-wide search for length polymorphic regions within genes, by comparing the JAY291 haplotype sequence to those of S288c and other sequenced *S. cerevisiae* strains. This approach allowed us to develop markers that can positively identify JAY270/PE-2 using a relatively small number of loci. We found that these new markers are also polymorphic in other major bioethanol strains, as well as in uncharacterized indigenous yeast contaminants. This targeted PCR-based monitoring of the microbial population is simpler, faster and more effective than the current methods employed by bioethanol distilleries.

METHODS

Strains and culture

The *S. cerevisiae* strains used in this study were predominantly derived from Brazilian bioethanol distilleries or *cachaça* still (listed in Supplementary Table 1; (ARGUESO *et al.* 2009; BASSO *et al.* 2008; CAMPOS *et al.* 2010; GOMES *et al.* 2000; WINSTON *et al.* 1995)). The main industrial strains used in sugarcane bioethanol production are of indigenous origin, and were all isolated as high productivity invaders. In addition, fermentation samples were collected from six different Brazilian bioethanol distilleries during the 2011 sugarcane harvest season to monitor the yeast population (Table 3). Samples collected at bioethanol production plants were streaked onto YPD solid medium [yeast extract 1% (w/v), peptone 2% (w/v), glucose 2% (w/v) and agar 2% (w/v)] for single colony isolation, and incubated at 30C for 2 days. Colonies were then cultured overnight in 4 ml YPD liquid medium at 30C for genomic DNA extraction.

Identification and selection of length polymorphic regions within genes

In order to identify polymorphic regions within genes, the coding regions from the *S. cerevisiae* reference genome (strain S288c) were aligned to the complete genome sequence of JAY291 using Exonerate (SLATER and BIRNEY 2005), applying stringent parameters to produce accurate sequence alignments (-best 1 -model est2genome -percent 20 -forcectag TRUE -quality 80 -refine region). A custom PERL script was created to parser the Exonerate output, identifying the insertion and deletion polymorphisms present in the output by alignment gaps. Only genes containing regions larger than 24 consecutive gaps and with gap lengths divisible by three (unaltered reading frames) were selected, resulting in an initial list of 27 candidate genes with usable length polymorphic regions. The sequences of these genes were then analyzed in the genomes of the unrelated *S. cerevisiae* strains YJM789 (WEI *et al.* 2007), AWRI1631 (BORNEMAN *et al.* 2008), and RM11-1a (*S. cerevisiae* RM11-1a Sequencing Project; <http://www.broad.mit.edu/>) through nucleotide similarity searches in GenBank using BLASTn. Of the 27 initial sequences analyzed, the eleven genes (Table 1) that displayed the highest degree of length variability between strains were selected as the most suitable for use as PCR marker loci.

Amplification of length polymorphic regions

Primer3 software (ROZEN and SKALETSKY 2000) was used to design oligonucleotide primers to amplify the selected polymorphic regions. In nine of the eleven loci the primers were

within the coding sequence. The exceptions were *YKL201C* in which the reverse primer was downstream of the coding sequence, and *YDR504C* with both primers located outside the coding region. The primer sequences used to amplify target regions are shown in Table 1.

Yeast genomic DNA was extracted using a phenol-chloroform protocol (AUSUBEL *et al.* 1998). PCR was carried out in 10 µl volume reactions containing 20 ng of genomic DNA, 0.25 mM dNTP (each), 1x PCR buffer, 3.5 mM MgCl₂, 0.05 U/µl of *Taq* DNA polymerase, 1.5 pmol/µl of each primer. The amplification program consisted of 5 minutes denaturation at 94C, followed by 40 cycles of 45 seconds at 94C, 40 seconds at primer-specific annealing temperature (Table 1), and 45 seconds at 72C. A 5 minute final extension was performed at 72C. PCR amplicons were resolved in conventional electrophoresis on 2% agarose gels in 1x TAE buffer (40mM Tris-acetate, 2mM Na₂EDTA.2H₂O, pH 8.5). Gels were stained with ethidium bromide, visualized under UV light, and photographed. The size and number of the detected bands was scored relative to a molecular weight ladder.

***In silico* PCR analysis**

The S288c sequences of the eleven selected loci, in their respective regions delimited by the primer pairs listed in Table 1 were used as reference for *in silico* PCR analysis. The comparison of sequence lengths at each locus in multiple *S. cerevisiae* strains was conducted by BLASTn searches using the complete genomic sequences deposited at the Sanger Institute, SGD and NCBI databases.

Re-sequencing of specific loci in JAY291

As part of this work, we re-sequenced the *YKL163W* and *YKL201C* loci from haploid strain JAY291 because they were found to be mis-assembled in the original JAY291 genome. The new correct sequences were used to edit the existing JAY291 contigs deposited in GenBank under accession numbers ACFL01000239.2 and ACFL01000006.2, respectively. The *YFL024C*, *YKL201C*, and *YLL021W* loci from the sibling haploid JAY292 were deposited under GenBank accession numbers JX101677, JX101678, and JX101679, respectively. All these sequences were obtained by conventional Sanger sequencing of PCR products generated with the primers in Table 1.

RESULTS and DISCUSSION

Identification of length polymorphic regions within genes

The aim of this study was to develop high specificity molecular markers for fingerprinting Brazilian bioethanol yeast strains based on the information derived from the JAY270/PE-2 strain genome sequence (ARGUESO *et al.* 2009). We compared coding sequences from the genome of the standard laboratory strain S288c (GOFFEAU *et al.* 1996) to those of strain JAY291, a haploid derivative of JAY270/PE-2. We then refined the search for length polymorphic loci by examining candidate regions in three other unrelated sequenced strains (YJM789, AWRI1631, and RM11-1a; METHODS). Eleven genes that contained sufficiently large continuous length polymorphic regions between these strains were identified (Table 1). In most cases, the polymorphic regions contained low complexity sequences with short repetitive patterns, often encoding sequential repetitions of codons for one or a few amino acid residues (Figure 1; and Supplementary Figure 1). Several of the identified genes had been previously reported to contain trinucleotide repeats within their coding regions (RICHARD *et al.* 1999; YOUNG *et al.* 2000). Others, such as *YLL021W*, had internal repeat units that were larger in size (81 bp). In all cases, the length polymorphic regions were present as in-frame insertions or deletions in the S288c and JAY291 coding sequences, and none of the selected genes contained introns. Four of the genes (*YDR299W*, *YFL024C*, *YJL005W*, and *YPR143W*) are essential for viability in S288c according to the *S. cerevisiae* systematic deletion studies. The eleven genes participate in assorted cellular pathways. The primary goal of this study was to use of these sequences as molecular strain identification markers, therefore we did not investigate any possible association between locus function and the presence of length polymorphisms.

Once the candidate length polymorphisms were identified, we designed oligonucleotide primer pairs flanking those sites to generate PCR products (Table 1). We used these primers to amplify genomic DNA from FY86 (an S288c isogenic strain; (WINSTON *et al.* 1995)) and from JAY291, and confirmed the expected product size polymorphisms in nine of the eleven genes (Figure 2A). For two of the genes, *YKL163W* and *YKL201C*, the resulting JAY291 PCR products had different sizes from those predicted from the published genomic sequence. However, in both cases the observed bands were still polymorphic relative to S288c. The likely cause for this discrepancy was local error in the *de novo* assembly of the JAY291 genome (ARGUESO *et al.* 2009). Repetitive and low complexity regions are particularly challenging to assemble accurately from the short sequencing reads used to determine the JAY291 genomic sequence (mostly 35 bp Illumina GAII reads, and some 200 bp Roche 454). Therefore, we re-sequenced the JAY291

PCR products spanning the *YKL163W* and *YKL201C* polymorphic regions using long reads from conventional Sanger sequencing. The Sanger-determined sequences of both genes matched the sizes of the PCR products (Figure 1, and Figure 2A).

PCR identification of industrial bioethanol strains

Following the discovery of the intragenic length polymorphisms between S288c and JAY291, we used the respective primer sets to look for polymorphisms between our strain of interest JAY270/PE-2, and strains BG-1, CAT-1 and SA-1 (Figure 2B). This group currently represents the four most widely adopted strains in the Brazilian bioethanol industry (BASSO *et al.* 2008). This analysis revealed polymorphisms at eight of the eleven loci assayed, confirming that our set of markers was highly informative for the identification bioethanol strains.

In several cases, more than one PCR product was amplified from genomic DNA of the industrial strains. We previously showed that JAY270/PE-2 is a highly heterozygous diploid, therefore suggesting that this strain, as well as the other industrial strains tested, may be heterozygous at the polymorphic loci. To test this hypothesis, we prepared genomic DNA from JAY270/PE-2 and from all four haploid spore products of a JAY270/PE-2 meiotic tetrad (Figure 3). We used this material as template in PCR reactions to amplify three candidate homozygous loci (*YDR299W*, *YDR505C*, and *YKL163W*) and three candidate heterozygous loci (*YFL024C*, *YKL201C*, and *YLL021W*). The four haploid spores in this tetrad included JAY291 (the one with a fully sequenced genome), and its three sibling spores JAY289, JAY290, and JAY292. For the homozygous markers, we observed a single PCR band of the same size in all samples. In contrast, for the *YFL024C* and *YLL021W* candidate heterozygous loci we observed the amplification of two bands in the JAY270/PE-2 diploid, and a single band in each of the haploid spores. PCR at the *YKL201C* locus produced three bands in JAY270/PE-2, the two shorter products corresponded to the alleles observed in the haploids, and the longer third band was a specific and reproducible chimera (characterization and discussion presented in Supplementary Figure 3). In addition, the polymorphic PCR products from the three heterozygous loci displayed the expected 2:2 Mendelian segregation among the spores. Finally, we sequenced the second (unknown) alleles from *YFL024C*, *YLL021W*, and *YKL201C*, and compared them to the first alleles already known from the JAY291 genomic sequence (Figure 1). In the case of *YFL024C*, the allele found in JAY290 and JAY292 was 385 bp long, and was the same as in S288c. For *YLL021W*, we observed that the allele present in JAY290 and JAY292 had two units of an 81 bp internal repeat, while the JAY291 allele has three units, and S288c has only one unit. Finally, in

the case of *YKL201C*, the allele found in JAY290 and JAY292 was 583 bp long, and differed from both the 631 bp allele in JAY291 and JAY289, and also from the 699 bp allele from S288c. This analysis showed that the high degree heterozygosis observed in JAY270/PE-2 and in other yeast strains is a helpful feature that further enhances the discriminatory power of our PCR markers.

To further assess the usefulness of the PCR markers, we analyzed a larger panel of yeast strains associated with sugarcane fermentation. These included the major strains described above, as well as a baker's yeast strain often used as a starter inoculum (FF), a *Cachaça* still strain (CA-11), three minor bioethanol strains (IZ1904, Y904, M304-2C), and eleven indigenous contaminant yeast strains isolated from Brazilian distilleries. The PCR products were visualized in agarose gel electrophoresis (Supplementary Figure 2), and the product sizes for each strain were examined to identify their specific allelic length variants. The summary of this analysis is shown in Table 2. We observed several new PCR alleles that were amplified in the different strain / loci combinations. In one case, *YJL005W*, the only size polymorphism observed was the one originally mapped between S288c and JAY291. All other loci produced between three and eleven PCR products of discernible size. Four of the loci, *YFL024C*, *YKL163W*, *YKL201C*, and *YLL021W*, were particularly polymorphic, and we found that using only those loci provided sufficient discriminatory power to uniquely identify each and every one of the 21 strains in Table 2. For the remainder of our study, we focused our PCR genotyping analyses on those four core markers.

Monitoring of the yeast cell population during industrial fermentations

After completing the design and testing of our high-specificity PCR markers, we demonstrated their effectiveness by monitoring the yeast microbial population in samples collected from six Brazilian distilleries during the 2011 sugarcane harvest season. Each of these distilleries began their fermentations in March or April using known specific mixes of commercial dry active yeast powder as initial inoculum (Table 3). As is typical in Brazil, the starting inoculum was often composed of a large proportion of baker's yeast strains (*i.e.* Angest, FF), supplemented with a smaller amount of bioethanol-specific strains such as PE-2, CAT-1 and BG-1. During the first few weeks of cell recycling the baker's strains gradually disappear, while the bioethanol-specific strains and/or indigenous contaminants progressively take over the population. This dynamic has been demonstrated previously and it is a common feature of the Brazilian bioethanol process (BASSO *et al.* 2008; DA SILVA *et al.* 2005).

We collected cell samples from each distillery roughly at the end of the first and second trimesters of the harvest season: June/July and September/October, respectively. We prepared genomic DNA from about twenty yeast colonies isolated from each sample, and used those DNAs as template in PCR. The amplification patterns from these colonies were then compared to those of the strains known to have been originally present in the inoculum. It is important to note that banding patterns in the industrial strain colonies re-isolated from the industry were identical to the pattern previously observed in the laboratory stocks. This suggested that the PCR markers were genetically stable through extensive cell propagation and fermentation cell growth.

The banding patterns for the colonies isolated from the distilleries were analyzed to generate a snapshot of the yeast population at each time (Table 3). Using only the four core markers, we were able quickly and positively classify the colonies as being either derived from one of the initial starter strains, or indigenous when their banding pattern did not match any of those present in the inoculum.

As expected, the starter baker's yeast strains were mostly outcompeted by the time of the first sampling, and were not detected at all in the second. The bioethanol strains had varying degrees of success in establishing themselves as the dominant members of the yeast population. For example, PE-2 was very effective in colonizing the fermentation environment at the Alcídia and Eldorado distilleries, but was replaced by better-adapted indigenous contaminants at Alto Taquari. Our PCR data showed that the indigenous yeasts were represented by multiple strains, ranging from two and up to eleven unique PCR genotypes depending on the distillery. Interestingly, at the Alto Taquari distillery a specific indigenous strain was detected at a high frequency in the yeast population (~60%), yet the overall bioethanol yield in the distillery remained satisfactory (data not shown).

***In silico* genotyping: Use of the PCR markers in non-sugarcane yeast strains**

Our analysis of a panel of sugarcane-associated strains (Table 2) revealed a high degree of polymorphism for many of our PCR markers. In addition, our analysis of yeast samples collected at distilleries identified further allelic variants among indigenous contaminant strains (Table 3; and data not shown). Together these observations suggested that these loci are highly variable in structure, and that their usefulness as molecular markers may not be limited just to sugarcane strains, and it may extend to *S. cerevisiae* in general. To assess the length variability at our eleven loci, we took advantage of the availability of several fully sequenced *S. cerevisiae*

genomes. We created BLASTn alignments using S288c as the reference sequence, and compared the sequence lengths at each locus in 59 strains from different origins (BORNEMAN *et al.* 2008; LITI *et al.* 2009; WEI *et al.* 2007). This analysis, summarized in Supplementary Table 2, allowed us to uniquely differentiate all but two of the strains tested. These two strains (UWOPS03.461.4 and UWOPS05.227.2) were isolated from the same geographic region, and are very similar at the genomic level (LITI *et al.* 2009).

For most loci, the sequence alignments revealed a higher number of length variant alleles than we were able to observe in the sugarcane strain set (Table 2, and Supplementary Figure 2). This number ranged from only two alleles in *YBL046W* and *YJL005W*, up to seventeen alleles in *YKL201C*; however, we cannot rule out the possibility that some of the *in silico* length polymorphisms are artifacts due to mis-assembly of short sequencing reads at the low complexity regions. We found cases in which the size difference between the predicted PCR products for specific alleles was quite large and easily detectable in conventional agarose electrophoresis gels (e.g. 240 bp, 192 bp, 164 bp), but also observed very small size differences (<15 bp), that would require running the PCR products on higher-resolution polyacrylamide or capillary gels to allow visualization of the polymorphisms. Despite these limitations, the *in silico* PCR analysis confirmed that the markers are highly polymorphic in strains outside the sugarcane set, and also that in many cases the PCR length polymorphisms should be detectable without the aid of a sophisticated electrophoresis apparatus. Therefore the PCR markers developed in this study should be useful also in most other fermentation processes that require a simple and fast method for monitoring the yeast population.

CONCLUSIONS

Previously described PCR-based methods for the identification of yeast strains relied primarily on DNA sequences predicted to be highly variable in the genome of the laboratory strain S288c. In contrast, the strategy we adopted in this study was to directly identify length polymorphisms in our strain of interest, and then design PCR markers around those regions known to be informative. In addition to conventional microsatellites, our analysis expanded the pool of polymorphic regions to include those with low complexity sequences that are not perfectly repeated (e.g. *YKL201C*) and regions with longer low copy repeats (e.g. *YLL021W*). This approach allowed us to create a simple, fast and highly reproducible system for positively identifying yeast colonies derived from the JAY270/PE-2 strain in the heterogeneous yeast populations found in bioethanol distilleries. In addition, we showed that this set was also highly

polymorphic among other strains associated with sugarcane, and in strains from other unrelated origins. The PCR-based method described here has at least three major advantages over PFGE, the current standard for yeast population monitoring in the industry: (1) The PCR product banding pattern is easier to interpret; (2) the DNA sample preparation is simpler, faster, and more affordable; (3), the signal from specific strains is unique and more stable over time. All of these desirable features make this set of molecular markers a promising alternative technology that should be readily adoptable by bioethanol and other fermentation industries.

LIST OF ABBREVIATIONS

Pulsed-field gel electrophoresis (PFGE), base pairs (bp), polymerase chain reaction (PCR).

COMPETING INTERESTS

The authors report no conflicts of interest.

AUTHOR'S CONTRIBUTIONS

All authors contributed to the data analysis, and read and approved the final manuscript. O.V.C.N designed the concept, performed experiments and wrote the manuscript; M.F.C. and G.G.L.C performed bioinformatics analysis to identify polymorphic regions; W.O.B. performed industrial genotyping analysis; J.L.A. designed the concept, performed experiments and wrote the manuscript; G.A.G.P. wrote the manuscript.

ACKNOWLEDGEMENTS

We would like to thank ETH Bioenergia S/A and Brazilian funding agencies FAPESP and CNPq for partial funding support to this project.

ENDNOTES

The use of the marker loci described in this work for commercial / for profit yeast genotyping purposes is subject to intellectual property protection through a pending patent application submitted to the Brazilian National Institute of Intellectual Property (Instituto Nacional de Propriedade Intelectual - INPI); patent application number 018100029693, 2010.

REFERENCES

- AMORIM, H. V., M. L. LOPES, J. V. DE CASTRO OLIVEIRA, M. S. BUCKERIDGE and G. H. GOLDMAN, 2011 Scientific challenges of bioethanol production in Brazil. *Applied Microbiology and Biotechnology* **91**: 1267-1275.
- ARGUESO, J. L., M. F. CARAZZOLLE, P. A. MIECZKOWSKI, F. M. DUARTE, O. V. C. NETTO *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Research* **19**: 2258-2270.
- AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN *et al.*, 1998 *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.
- BASSO, L. C., H. V. DE AMORIM, A. J. DE OLIVEIRA and M. L. LOPES, 2008 Yeast selection for fuel ethanol production in Brazil. *Fems Yeast Research* **8**: 1155-1163.
- BORNEMAN, A. R., A. H. FORGAN, I. S. PRETORIUS and P. J. CHAMBERS, 2008 Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Research* **8**: 1185-1195.
- BOVO, B., C. ANDRIGHETTO, M. CARLOT, V. CORICH, A. LOMBARDI *et al.*, 2009 Yeast population dynamics during pilot-scale storage of grape marcs for the production of Grappa, a traditional Italian alcoholic beverage. *International Journal of Food Microbiology* **129**: 221-228.
- CAMPOS, C. R., C. F. SILVA, D. R. DIAS, L. C. BASSO, H. V. AMORIM *et al.*, 2010 Features of *Saccharomyces cerevisiae* as a culture starter for the production of the distilled sugar cane beverage, cachaca in Brazil. *Journal of Applied Microbiology* **108**: 1871-1879.
- CARRO, D., and B. PINA, 2001 Genetic analysis of the karyotype instability in natural wine yeast strains. *Yeast* **18**: 1457-1470.
- DA SILVA, E. A., S. K. B. DOS SANTOS, A. D. M. RESENDE, J. O. F. DE MORAIS, M. A. DE MORAIS *et al.*, 2005 Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **88**: 13-23.
- EZOV, T. K., E. BOGER-NADJAR, Z. FRENKEL, I. KATSPEROVSKI, S. KEMENY *et al.*, 2006 Molecular-genetic biodiversity in a natural population of the yeast *Saccharomyces cerevisiae* from "Evolution canyon": Microsatellite polymorphism, ploidy and controversial sexual status. *Genetics* **174**: 1455-1468.
- GOFFEAU, A., B. G. BARRELL, H. BUSSEY, R. W. DAVIS, B. DUJON *et al.*, 1996 Life with 6000

- genes. *Science* **274**: 546, 563-547.
- GOMES, L. H., K. M. R. DUARTE, J. L. ARGUESO, S. ECHEVERRIGARAY and F. C. A. TAVARES, 2000 Methods for yeast characterization from industrial products. *Food Microbiology* **17**: 217-223.
- HENNEQUIN, C., A. THIERRY, G. F. RICHARD, G. LECOINTRE, H. V. NGUYEN *et al.*, 2001 Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains. *Journal of Clinical Microbiology* **39**: 551-559.
- HIERRO, N., B. ESTEVE-ZARZOSO, A. GONZALEZ, A. MAS and J. M. GUILLAMON, 2006 Real-time quantitative PCR (QPCR) and reverse transcription-QPCR for detection and enumeration of total yeasts in wine. *Applied and Environmental Microbiology* **72**: 7148-7155.
- KURTZMAN, C. P., and C. J. ROBBETT, 1998 Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Van Leeuwenhoek* **73**: 331-371.
- LEGRAS, J. L., and F. KARST, 2003 Optimisation of interdelta analysis for *Saccharomyces cerevisiae* strain characterisation. *FEMS Microbiology Letters* **221**: 249-255.
- LEGRAS, J. L., O. RUH, D. MERDINOGLU and F. KARST, 2005 Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. *International Journal of Food Microbiology* **102**: 73-83.
- LITI, G., D. M. CARTER, A. M. MOSES, J. WARRINGER, L. PARTS *et al.*, 2009 Population genomics of domestic and wild yeasts. *Nature* **458**: 337-341.
- MARTORELL, P., A. QUEROL and M. T. FERNANDEZ-ESPINAR, 2005 Rapid identification and enumeration of *Saccharomyces cerevisiae* cells in wine by real-time PCR. *Applied and Environmental Microbiology* **71**: 6823-6830.
- MERCADO, L., S. JUBANY, C. GAGGERO, R. W. MASUELLI and M. COMBINA, 2010 Molecular relationships between *Saccharomyces cerevisiae* strains involved in winemaking from Mendoza, Argentina. *Current Microbiology* **61**: 506-514.
- NADAL, D., D. CARRO, J. FERNANDEZ-LARREA and B. PINA, 1999 Analysis and dynamics of the chromosomal complements of wild sparkling-wine yeast strains. *Applied and Environmental Microbiology* **65**: 1688-1695.
- PEREZ, M. A., F. J. GALLEGU, I. MARTINEZ and P. HIDALGO, 2001 Detection, distribution and selection of microsatellites (SSRs) in the genome of the yeast *Saccharomyces cerevisiae* as molecular markers. *Letters in Applied Microbiology* **33**: 461-466.
- QUEROL, A., E. BARRIO, T. HUERTA and D. RAMON, 1992 Molecular monitoring of wine

- fermentations conducted by active dry yeast strains. *Applied and Environmental Microbiology* **58**: 2948-2953.
- RICHARD, G. F., C. HENNEQUIN, A. THIERRY and B. DUJON, 1999 Trinucleotide repeats and other microsatellites in yeasts. *Research in Microbiology* **150**: 589-602.
- ROZEN, S., and H. J. SKALETZKY, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 365-386 in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, edited by S. M. KRAWETZ, S. Humana Press, Totowa, NJ.
- SCHULLER, D., L. PEREIRA, H. ALVES, B. CAMBON, S. DEQUIN *et al.*, 2007 Genetic characterization of commercial *Saccharomyces cerevisiae* isolates recovered from vineyard environments. *Yeast* **24**: 625-636.
- SCHULLER, D., E. VALERO, S. DEQUIN and M. CASAL, 2004 Survey of molecular methods for the typing of wine yeast strains. *FEMS Microbiology Letters* **231**: 19-26.
- SLATER, G. S., and E. BIRNEY, 2005 Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**.
- TECHERA, A. G., S. JUBANY, F. M. CARRAU and C. GAGGERO, 2001 Differentiation of industrial wine yeast strains using microsatellite markers. *Letters in Applied Microbiology* **33**: 71-75.
- WEI, W., J. H. MCCUSKER, R. W. HYMAN, T. JONES, Y. NING *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 12825-12830.
- WHEALS, A. E., L. C. BASSO, D. M. G. ALVES and H. V. AMORIM, 1999 Fuel ethanol after 25 years. *Trends in Biotechnology* **17**: 482-487.
- WINSTON, F., C. DOLLARD and S. L. RICUPEROHOVASSE, 1995 Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288c. *Yeast* **11**: 53-55.
- YOUNG, E. T., J. S. SLOAN and K. VAN RIPER, 2000 Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics* **154**: 1053-1068.

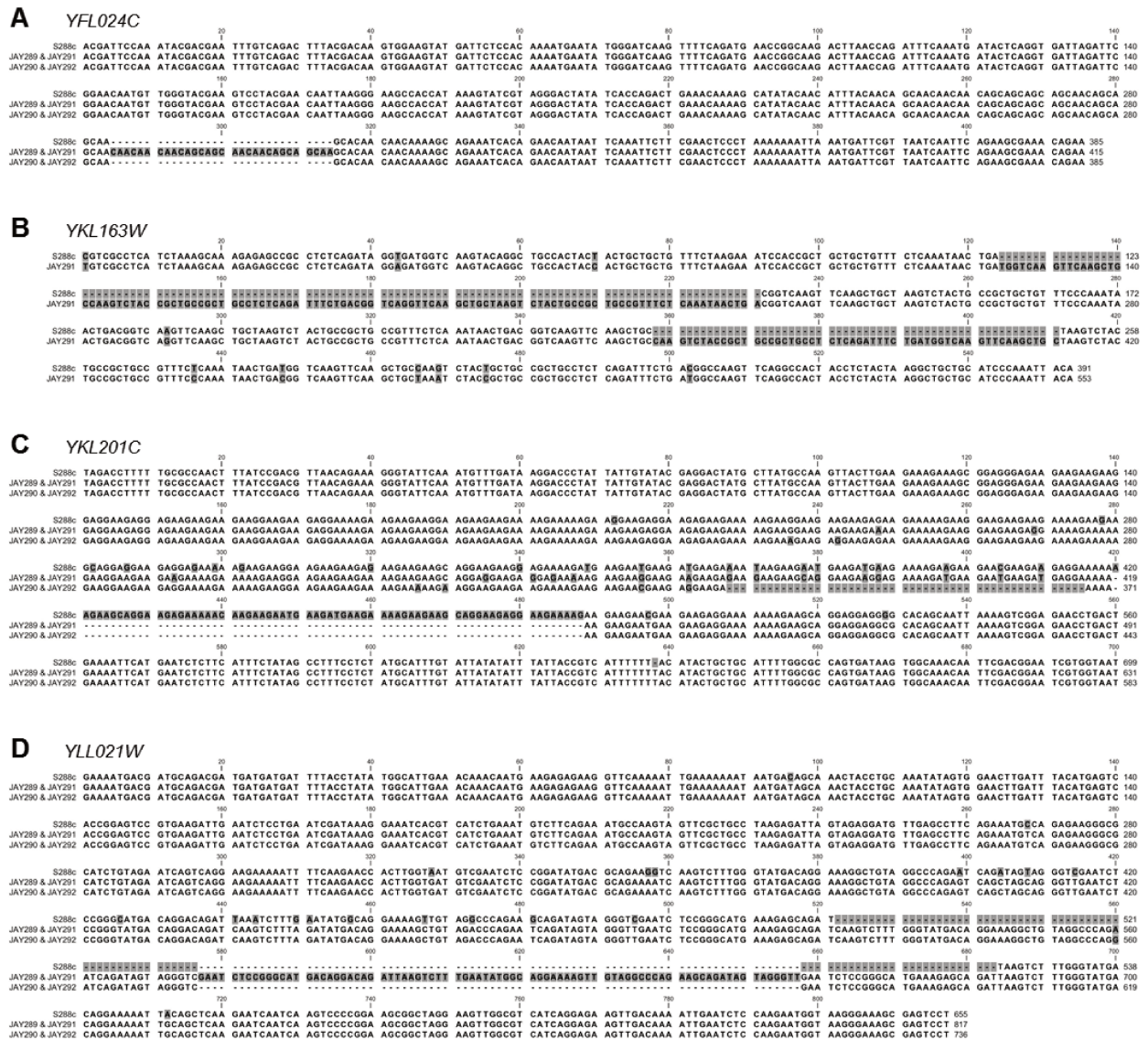


Figure 1. Length polymorphisms within coding sequences.

Alignments of four of the genomic DNA sequences used as PCR marker loci in this study. Panels A through D show the alignments between the predicted PCR products (from Forward to Reverse primers) from the reference strain S288c and from JAY270/PE-2 sequences. Polymorphisms are shown shaded in gray, and insertion/deletion gaps are shown as dashes. For the loci where JAY270/PE-2 is homozygous (B) only the JAY291 haploid sequence is shown; and for the loci where this strain is heterozygous (A, C, D), the sequences of the two alleles are shown separately.

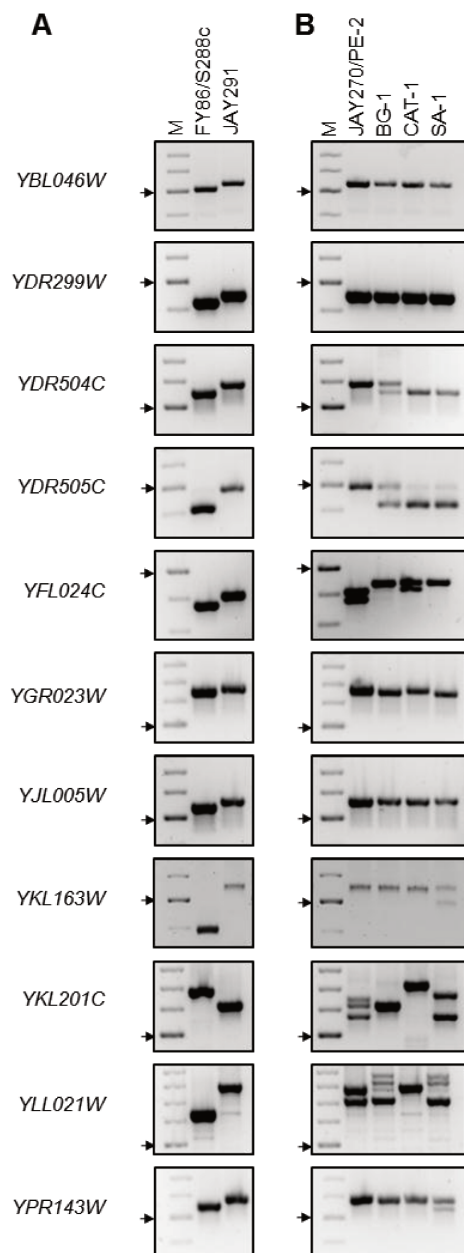


Figure 2. PCR amplification of length-polymorphic loci.

Panel A shows agarose gels with PCR products amplified from the genomes of the reference strain FY86/S288c and of the JAY291 haploid. Panel B shows the PCR products obtained from the JAY270/PE-2 heterozygous diploid, and from three other widely used bioethanol yeast strains. M indicates the molecular weight ladder, and the arrow indicates the 500 bp marker band. The only exception is for the *YLL021W* gel for which the 700 bp band is indicated.

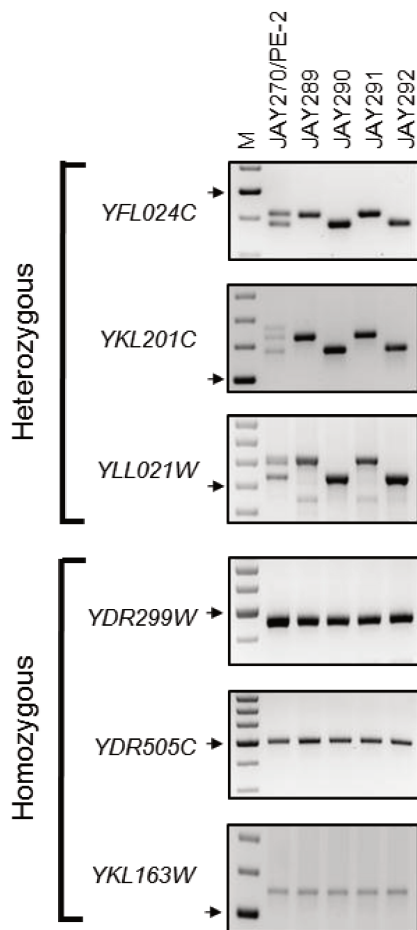


Figure 3. Analysis of heterozygous length polymorphisms in JAY270/PE-2

Agarose gels showing the PCR products amplified from the genome of the JAY270/PE-2 heterozygous diploid, and from the four haploid spores JAY289, JAY290, JAY291, and JAY292. These four spores are derived from a single tetrad of JAY270. The top panel shows polymorphic markers for three heterozygous loci, with 2:2 Mendelian segregation among the haploids. The bottom panel shows the banding patterns for three homozygous loci. M indicates the molecular weight ladder, and the arrow indicates the 500 bp marker band. The only exception is for the *YLL021W* gel for which the 700 bp band is indicated.

Table 1. List of genes, PCR products, and primers flanking the length polymorphic regions.

ORF identifier	Gene name	PCR product size (bp)		5'-3' primer sequences		Melting temperature (C)
		S288c	JAY291	Forward	Reverse	
<i>YBL046W</i>	<i>PSY4</i>	514	541	GCAGTTGGCAGGTTTTTGAT	GCCTGCCTATCTTCGTCATC	55
<i>YDR299W</i>	<i>BFR2</i>	433	460	AGCGGATCAAATTTCCGATA	TGCATGCCTTTTCTGTTGAG	52
<i>YDR504C</i>	<i>SPG3</i>	564	600	TCGACAACTTTCAGATTGCTG	AATGGAAGCGGCTGTTAGAA	55
<i>YDR505C</i>	<i>PSP1</i>	425	497	ACAGTTTCGCCGGAATACAC	CCATGTTCTTCGCGTCATAA	52
<i>YFL024C</i>	<i>EPL1</i>	385	415	ACGATTCCAAATACGACGAA	TTCTGTTTCGCTTCTGAATTG	57
<i>YGR023W</i>	<i>MTL1</i>	664	679	AAGGCCAGACCATCCTCTCT	CCGTTGTTTGGTTTCGTGTA	53
<i>YJL005W</i>	<i>CYR1</i>	550	574	CTCTGACAACAAACGCGAAG	ATCGAGTGCGGTCTTGCTAT	57
<i>YKL163W</i>	<i>PIR3</i>	391	553 ^a	TGTCGCCTCATCTAAAGCAA	TGTAATTTGGGATGCAGCAG	54
<i>YKL201C</i>	<i>MNN4</i>	699	631 ^b	TAGACCTTTTTGCGCCAACT	ATTACCACGATTCCGTCGAA	57
<i>YLL021W</i>	<i>SPA2</i>	655	817	GAAAATGACGATGCAGACGA	AGGACTCGCTTTCCTTACC	55
<i>YPR143W</i>	<i>RRP15</i>	562	601	GGGCTCCAAGCACAGAGTAG	CTTTCTGAGCGCTGTCTCCT	57

a. Re-sequenced PCR product; The predicted PCR product size for *YKL163W* according to the first JAY291 genomic assembly was 337 bp.

b. Re-sequenced PCR product; The predicted PCR product size for *YKL201C* according to the first JAY291 genomic assembly was 582 bp.

Table 2. Analysis of PCR length polymorphic alleles in common sugarcane yeast strains.

Strains:	YBL046W	YDR299W	YDR504C	YDR505C	YFL024C	YGR023W	YJL005W	YKL163W	YKL201C	YLL021W	YPR143W
JAY270/PE-2	2	2	2	2	1/2	2	2	2	2/3/6	2/9	2
CAT-1	2	2	1	1	3	3	2	2	4	4	2
BG-1	2	2	1/2	1/2	3	1	2	2	6	3/8/10	2
SA-1	2	2	1	1	3	1	2	2/7	2/5	3/8/10	1/2
FY86/S288c	1	1	1	1	1	1	1	1	1	1	1
CA-11	2	2	3	1	3	1	2	7	2	1	1
IZ1904	2	2	1	3	4	2	2	2	5	5	1
Y904	1	3/4	1	1	1	1	2	8	1	3	2
M304-2C	1	1	1	1	1	1	2	8	1	1	1
FF	2	3	1	1	2	2/3	2	7/4	5/7	4/3	2
CL-A	2	1/2	1	1	2/5	3	2	4	6/7	4	3
SJC1 230/8A	2	3	1	1	4	3	2	2/7	2	2/9	1
SJC1	2	3	2	4	6	1	2	3	8/11	5	1
SJC3	2	3	1	1	2/6	1	2	2	11	2/9	1
SJC7	2	3	1	1	4	2/3	2	2/7	2/6/11	4	1/2
PRGC4A	2	5	1	1	4/6	2/3	2	2	3/9	6/7/10	1/4
EQ287C1	1/2	3	1	1	2	3	2	2/7	4/9	6/7	1
EQ287C2	2	2	1	1	2	1	2	5/6	10	7	1/4
LQ1	2	2	1	1	2/4	2	2	2	1	7	4
LQ2	2	2/6	1	1	4/6	2/3	2	2	1/4/9	3/7/10	1/2
SJC6	2	2	3	1	4/6	2/3	2	2/7/9	6/9/10	6/7	1

The numbers identify the specific PCR allele detected in each strain. The allele observed in the FY86/S288c reference genome was arbitrarily assigned as allele "1". The allele observed in JAY291 (Figure 1A) was arbitrarily assigned as allele "2".

Table 3. PCR monitoring of yeast populations in six bioethanol distilleries during the 2011 season.

Strain:	Distilleries																	
	Alcídia			Morro Vermelho			Eldorado			C. do Pontal			Santa Luzia			Alto Taquari		
	^a Inoc. (tons)	^b 1 st	^c 2 nd	Inoc. (tons)	1 st	2 nd	Inoc. (tons)	1 st	2 nd	Inoc. (tons)	1 st	2 nd	Inoc. (tons)	1 st	2 nd	Inoc. (tons)	1 st	2 nd
PE-2	0.35	22	18	0	0	0	0.1	12	19	0	0	0	0.3	0	0	10	10	0
CAT-1	0.35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	5	0
BG-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	2	0
Angest	0	0	0	4	0	0	1.5	0	0	2	1	0	2	0	0	0	0	0
FF	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
^d Indigenous	0	13	2	0	15	15	0	3	1	0	14	20	0	20	20	0	3	20
Total	0.7	35	20	4	15	15	1.6	15	20	2	15	20	4.3	20	20	30	20	20

This analysis only included the molecular markers for loci *YFL024C*, *YKL163W*, *YKL201C* and *YLL021W*. Colonies with a heterozygous, or a homozygous pattern for one of the two *YFL024C*, *YKL201C* or *YLL021W* alleles present in JAY270/PE-2 were classified as derived from the starter strains.

a. Tons of dry active yeast powder used as starting inoculum (March/April).

b. Number of colonies identified at the first sample collection after the beginning of the season (June/July).

c. Number of colonies identified at the second sample collection (September/October).

d. Colonies whose genotype that did not match any of the other industrial strains in the starting inoculum were classified as indigenous.

CAPÍTULO 3

“Processo para produção de uma levedura mutante nulo do gene *URA3* da cepa industrial Pedra-2 (*Saccharomyces cerevisiae*), levedura mutante e uso da mesma”

Felipe Galzerani; Osmar V. Carvalho-Netto; Juan Lucas Argueso; Fabiana M. Duarte; Sílvia K. Missawa; Gonçalo A. G. Pereira.

Pedido de Patente de Processo: ETH Bioenergia/UNICAMP, protocolada no INPI sob No. 018100039982, em 25/10/2010.

Processo para produção de uma levedura mutante nulo do gene *URA3* da cepa industrial Pedra-2 (*Saccharomyces cerevisiae*), levedura mutante e uso da mesma

Campo da Invenção

A presente invenção está relacionada ao campo da engenharia genética e biotecnologia. Mais especificamente, a presente invenção se refere a uma linhagem de levedura mutante *Saccharomyces cerevisiae* auxotrófica para uracila.

Antecedentes da Invenção

Tendo em vista as mudanças climáticas observadas nos últimos anos decorrentes do aquecimento global, no início da década de 90 foi estabelecido o Protocolo de Kyoto, a partir do qual as nações envolvidas se comprometeram a reduzir suas emissões de dióxido de carbono e outros gases causadores do efeito estufa, a níveis 5% inferiores até o ano de 2012, através da substituição de combustíveis fósseis derivados de petróleo por fontes renováveis de energia.

Como resultado desse comprometimento, o etanol tem se destacado nos últimos 30 anos como uma das principais fontes renováveis de energia, sendo hoje um biocombustível altamente competitivo frente aos oriundos do petróleo. Previsões indicam a possibilidade de o etanol substituir mais de 10% da gasolina mundial nos próximos 15 a 20 anos, de acordo com Goldemberg, 2008, *Biotechnol Biofuels*, 1(1): 6).

Grande parte deste mérito se deve ao Brasil e Estados Unidos que são os dois maiores produtores mundiais deste importante recurso energético. Embora os Estados Unidos detenham hoje a primeira posição no ranking de produtividade, o Brasil foi durante décadas o maior produtor de etanol do mundo, tendo sido superado pelos norte-americanos apenas no final de 2005 (RFA, 2008).

A supremacia brasileira nas décadas de 80 e 90 e, início do século XXI, se deve em grande parte ao Programa Nacional de Álcool (Proálcool), criado em 1975, que foi uma iniciativa do governo brasileiro em investir na produção de etanol como uma solução frente à elevação do preço do petróleo evidenciada em meados da década de 70. Assim sendo, o atual destaque do país na economia mundial é fruto de enormes investimentos que foram destinados ao setor sucroalcooleiro, que levaram a melhorias na agricultura, através de programas de melhoramento genético da própria matéria-prima (cana-de-açúcar) e do desenvolvimento de tecnologias de produção, distribuição e utilização do etanol de acordo com o trabalho apresentado por Goldemberg e colaboradores no ano de 2004, intitulado: "*Biomass and*

Bioenergy", 26(3): 301-304.

No Brasil, a produção de etanol ocorre por meio da via fermentativa nas dornas das usinas sucroalcooleiras, onde são adicionados o mosto (caldo de cana-de-açúcar) e as leveduras, as quais são responsáveis pela conversão da glicose presente na matéria-prima no produto final. Após o término da fermentação, todo o conteúdo é transferido para as centrífugas. O vinho, fermentado de cana sem leveduras, segue para as torres de destilação para a produção do etanol. O leite de leveduras, nome dado ao caldo de leveduras que sai da centrifugação, passa por um tratamento com ácido sulfúrico antes de retornar à dorna, na qual será adicionada uma nova carga de mosto para se ter início a um novo ciclo. Com o uso de centrífugas, cerca de 90% das células são reaproveitadas de uma fermentação para outra, sendo que o reciclo permanece durante toda a safra, de 200 a 250 dias conforme descrito por Meleiro & Maciel, 2000, *Braz. J. Chem. Eng.*, intitulado: "*State and parameter estimation based on a nonlinear filter applied to an industrial process control of ethanol production*", 17(4-7): 991-1001).

Uma constatação recente e de extrema importância para o setor é que em muitas usinas a população de linhagens de leveduras não se mostra homogênea e nem constante durante o processo fermentativo (Silva-Filho *et al.*, 2005, Antonie Van Leeuwenhoek, intitulado "Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting", 88: 13-23).

As leveduras inoculadas em altas concentrações no início da safra são substituídas em poucos dias por leveduras selvagens que acabam dominando o processo de fermentação (Silva-Filho *et al.*, 2005, Antonie van Leeuwenhoek, intitulado "Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting", 88: 13-23; Basso *et al.*, 2008, *FEMS Yeast Res.*, intitulado "Yeast selection for fuel ethanol production in Brazil", 8(7): 1155-63), onde no decorrer do processo fermentativo, as leveduras constantemente se deparam com diversos fatores de estresse, tais como alta concentração de etanol, pressão osmótica devido às altas concentrações de açúcar e sais, acidez, contaminação bacteriana, os quais geram um ambiente de competição que favorecem o desenvolvimento de cepas mais resistentes e melhor adaptadas, que muitas vezes vêm junto com a colheita da cana-de-açúcar (Alves, 1994, intitulado "Fatores que afetam a formação de ácidos orgânicos bem como outros parâmetros da fermentação alcoólica", MS Thesis).

Sabendo da existência desta diversidade microbiológica de leveduras atuando nas diferentes destilarias brasileiras, cepas selvagens de bom desempenho fermentativo e de ótima

adaptação às condições industriais das dornas vêm sendo selecionadas e isoladas a partir de diversas usinas.

As leveduras selvagens que mostraram resistência às condições de estresse e características fermentativas desejáveis foram reintroduzidas em inúmeras destilarias brasileiras para que fosse possível observar a sua capacidade de dominância e resistência em outras usinas frente às leveduras iniciadoras do processo e também frente às nativas das suas próprias localidades (Basso *et al.*, 2008, FEMS Yeast Res., intitulado “Yeast selection for fuel ethanol production in Brazil”, 8(7): 1155-63).

Dentre as leveduras selecionadas, se destacam as linhagens de *Saccharomyces cerevisiae* denominadas PE-2 (Pedra-2) e CAT-1 que foram justamente as mais utilizadas na safra de 2007 a 2008, por cerca de 150 usinas do Estado de São Paulo e, respondendo por 60% do etanol produzido no Brasil (Basso *et al.*, 2008, FEMS Yeast Res., intitulado “Yeast selection for fuel ethanol production in Brazil”, 8(7): 1155-63).

Atualmente a Pedra-2 é a cepa mais importante do país, sendo responsável por quase 40% da produção de etanol nacional e 10% de etanol mundial, o que tem despertado o interesse de alguns pesquisadores, os quais têm buscado compreender melhor os motivos que a torna uma excelente levedura de fermentação em escala industrial (Argueso *et al.*, 2009, Genome Research, intitulado “Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production”, 19(12): 2258-70; Stambuk *et al.*, 2009, Genome Research, intitulado “Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis”, 19(12): 2271-78).

Argueso e colaboradores, 2009, Genome Research, intitulado “Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production”, estão trabalhando com um isolado da cepa Pedra-2, denominado de linhagem JAY270. Os autores mostraram recentemente que a linhagem JAY270 é uma levedura diplóide geneticamente compatível com linhagens *Saccharomyces cerevisiae* de laboratório, como a S288c. Análises moleculares revelaram que a JAY270 possui um genoma altamente heterozigoto, tanto estruturalmente quanto em nível de nucleotídeos. De acordo com os autores, o sucesso da linhagem no ambiente altamente competitivo da dorna está intrinsecamente relacionado com a sua arquitetura genômica heterogênea. A presença de rearranjos nas partes periféricas dos seus cromossomos aliada ao alto grau de polimorfismo em nível de nucleotídeo permitiram que a linhagem se adaptasse bem as condições estressantes da dorna, mostrando um ótimo desempenho fermentativo. Grande parte destas informações emergiu do seqüenciamento e

montagem do genoma da JAY291, que é um dos quatro esporos (JAY289, JAY290, JAY291 e JAY292) de uma tétrade da levedura diplóide. Os autores afirmam, ainda, que a JAY270 parece ser uma linhagem de levedura híbrida formada pelo cruzamento natural de linhagens parentais distantes relacionadas (Argueso & Pereira, 2010, *International Sugar Journal*, 112: 86-89).

O recente trabalho de Argueso e colaboradores, 2009, *Genome Research*, intitulado “Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production”, 19(12): 2258-70 deixa claro que a linhagem JAY270 apresenta um enorme potencial de aplicação para o setor industrial. Se a Pedra-2 já tem mostrado, naturalmente, uma excelente capacidade de fermentação em escala industrial, não há dúvidas de que estudos envolvendo o melhoramento genético da linhagem JAY270 poderiam ser aplicados para aumentar a produção de etanol combustível e também para a confecção de outros produtos de interesse industrial, como glicerol e hidrocarbonetos que atualmente são derivados do petróleo. Considerando que o Brasil obteve uma produção de etanol na ordem de 25 bilhões de litros em 2008, um aumento ínfimo de 1% na capacidade de produção da Pedra-2, através de engenharia genética, representaria um aumento extraordinário de 250 milhões de litros.

Aliás, o melhoramento genético de linhagens de leveduras selvagens é uma área, ainda, pouquíssima estudada e que deve ser mais explorada, até mesmo considerando que desde a implementação do Proálcool os mais variados segmentos do setor sucroalcooleiro têm sido aperfeiçoados e incrementados com inovações tecnológicas, mas que na verdade, quase nada foi feito para compreender e melhorar geneticamente os agentes microbiológicos do processo de fermentação das usinas que são as leveduras industriais.

Embora esse seja um campo de estudo com grande potencial para o setor industrial, existem, até o momento, pouquíssimos trabalhos relacionados ao melhoramento genético de leveduras, principalmente, no que diz respeito ao de linhagens industriais.

Um dos motivos que dificulta o melhoramento genético de leveduras industriais é a falta de conhecimento sobre a natureza biológica e genética das suas linhagens, o que acaba inviabilizando a aplicação das diversas ferramentas moleculares de manipulação genética disponíveis atualmente (Sherman, 1998, *The encyclopedia of molecular biology and molecular medicine*, 6: 302-325).

Em contraste, leveduras de laboratório, tal como a linhagem de *Saccharomyces cerevisiae* S288c, têm sido amplamente estudadas nos últimos anos, de modo que elas estão muito bem descritas do ponto de vista biológico e genético, o que acaba facilitando e

viabilizando quaisquer estudos de manipulação genética envolvendo o melhoramento das suas linhagens. É válido lembrar neste ponto, que embora as leveduras de laboratório sejam importantes como modelo de pesquisa científica, sendo até mesmo alvo de estudos sobre características fermentativas, elas não são apropriadas nem mesmo naturais do ambiente fermentativo, de modo que não são empregadas nos processos industriais.

As leveduras são mundialmente reconhecidas pelo enorme potencial para os processos fermentativos industriais. Isto se deve as suas propriedades de rusticidade e robustez - como a habilidade de crescer anaerobicamente, tolerância ao pH baixo, a alta concentração de açúcar e etanol - que são exigidas durante o processo de fermentação em escala industrial. Durante anos, inúmeros estudos têm visado o melhoramento destas características na busca de leveduras melhor adaptadas às condições do processo de fermentação (Nevoigt, 2008, *Microbiology and Molecular Biology Reviews*, 72(3): 379-412).

Há estudos que visam à incorporação de caracteres que permitam a sobrevivência ao ambiente altamente estressante das dornas, tal como alta tolerância ao etanol e às altas concentrações de açúcares (Dequin, 2001, *Appl. Microbiol. Biotechnol.*, 56(5-6): 577-88; Jiménez-Martí *et al.*, 2009, *International Journal of Food Microbiology*, 130: 122-130). Outros têm buscado a obtenção de linhagens mais efetivas na produção de etanol (Teunissen & Steensma, 1995, *Yeast*, 11: 1001-1013; Cunha *et al.*, 2006, *Yeast Res.*, 6(2): 280-7). Jiménez-Martí e colaboradores (2009) recentemente conseguiram bons resultados ao super expressarem por meio de plasmídeos, genes de resistência ao estresse fermentativo em linhagens *S. cerevisiae* de vinificação.

Durante décadas, leveduras de laboratório da espécie *Saccharomyces cerevisiae*, como por exemplo, a linhagem S288c, passaram por modificações em genes marcadores envolvidos em vias metabólicas de biossíntese de aminoácidos, tais como triptofano, leucina e histidina, que ajudaram a compreender as suas funções.

Através de métodos nem um pouco específicos, tais como raios ultravioletas, agentes químicos, entre outras mutações em genes marcadores foram estudadas. Como resultado leveduras que levavam consigo mutações em um gene marcador específico passavam a ser incapazes de sintetizar o aminoácido correspondente, sendo consideradas auxotróficas para tal gene ou via metabólica.

A produção de mutantes auxotróficos, como pela inativação de genes envolvidos na síntese de certo aminoácido, permite a seleção de transformantes contendo o gene selvagem em meio carente no aminoácido em questão é descrito por Wery e colaboradores no Manual

of Industrial Microbiology Biotechnology, páginas 447-59 de 1999.

A aplicação de inúmeras ferramentas moleculares de manipulação genética também é descrito em diversos trabalhos, tais como, por Ausubel *e colaboradores* no Current Protocols in Molecular Biology em 1998, por Sherman no The encyclopedia of molecular biology and molecular medicine, 6: 302-325 de 1998.

Um dos principais genes marcadores de leveduras é o *URA3* que está envolvido na via de biossíntese de uracila. O gene *URA3* codifica a enzima orotidina-5'-fosfato (OMP) descarboxilase, responsável pela conversão de orotidilato em uridilato (UMP).

A inativação deste gene permite a contra-seleção dos mutantes em 5-fluoro-orótico (5-FOA), um composto estruturalmente análogo ao orotato, e que pode entrar na via de biossíntese da uracila gerando um produto tóxico (5-fluoroacil).

Assim, mutantes para o gene *URA3* (genótipo *ura3*) podem ser selecionados em meio suplementado com uracila e acrescentado de 5-FOA, já que somente os mutantes que não tenham o gene *URA3* ativo serão capazes de crescer em meio com 5-FOA como descrito por Boeke e colaboradores, 1984, Molecular and General Genetics, no trabalho intitulado: A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeasts: 5-fluoro-orotic acid resistance, 197: 345-346.

A obtenção deste tipo de mutante permite uma avaliação fenotípica rápida dos mutantes positivos, utilizando a complementação da via de biossíntese de uracila como marca de seleção auxotrófica.

Embora mutantes *ura3* sejam incapazes de crescer em meio sem uracila, ao adquirirem o gene selvagem *URA3* por transformação genética, eles passam a ser capazes de crescer no mesmo meio seletivo, de tal forma que uma modificação genética de interesse pode ser “rastreada” e, dessa forma, uma levedura geneticamente melhorada pode ser selecionada.

Devido à facilidade em se trabalhar com leveduras auxotróficas em estudos de manipulação genética, muitas leveduras foram modificadas no decorrer do tempo, para que genes marcadores pudessem ser utilizados como marca de seleção.

Consequentemente, hoje existem muitas leveduras de laboratório que tiveram genes marcadores interrompidos por sequências de DNA carregando consigo marcas exógenas (http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html), ou mesmo que sofreram mutações pontuais decorrentes de irradiação ultravioleta e que, com isso, passaram a apresentar a auxotrofia de interesse. Entretanto, a utilização de genes marcadores não é a única forma de se conseguir selecionar uma levedura transformante.

A seleção de leveduras transformantes, ou seja, daquelas com a modificação genética de interesse, pode ser feita com genes bacterianos de resistência a drogas e também com genes marcadores da própria natureza genética do microrganismo a ser manipulado. No primeiro caso, a levedura modificada passa a ser considerada um organismo transgênico, pela presença de um gene que não é da sua natureza genética, acarretando inúmeras complicações de ordem pública quanto a sua utilização para fins comerciais (Hug, 2008, Medicina, intitulada “Genetically modified organisms: do the benefits outweigh the risks?”, 44(2): 87-99).

Além disso, de acordo com as normas de biossegurança impostas pela Comissão Técnica nacional de Biossegurança (CTNBio), empresas que trabalham com um organismo geneticamente modificado que é transgênico precisam tomar medidas cabíveis a fim de controlar e evitar a sua disseminação no meio ambiente, o que representa despesas indesejáveis em termos financeiros. Assim sendo, pensando na utilização de um organismo geneticamente modificado para fins comerciais, o segundo caso acaba sendo mais interessante para as empresas. No entanto, para que seja viável o uso de técnicas de manipulação genética que não envolvam marcas de transgenia, mas sim genes marcadores, é preciso trabalhar com leveduras que apresentem auxotrofia para os mesmos e, assim, para as suas vias de biossíntese relacionadas.

A maneira como a modificação genética seria realizada na linhagem JAY270, para se conseguir a auxotrofia para uracila, foi um ponto crítico da presente invenção. Certamente teriam que ser utilizados métodos mais específicos e menos antiquados comparados aos que foram usados com as linhagens *S. cerevisiae* de laboratório, como a S288c, e também, comparados aos que, ainda, têm sido utilizados para a obtenção de leveduras industriais auxotróficas. Tais métodos contam, por exemplo, com o uso de irradiação ultravioleta e agentes químicos mutagênicos. Tendo em vista que a intenção mais adiante será melhorar geneticamente a JAY270 não seria cabível utilizar irradiação ultravioleta para gerar a auxotrofia para uracila, pois sendo este um método muito invasivo, ele é danoso ao próprio genoma da levedura e, assim, às suas características de interesse fermentativo, que dica de passagem devem ser totalmente preservadas. Prova disso é o fato de que, quando as leveduras de laboratório eram submetidas aos diversos procedimentos de manipulação – inclusive os que empregavam raios ultravioletas - para a inativação de genes, elas acumulavam mutações deletérias e acabavam perdendo muitas das suas características naturais encontradas nos seus ancestrais selvagens. Embora isso não tenha sido um grande problema para as leveduras de laboratório, já que elas são apenas veículos de estudos científicos, era algo que não poderia

ser repetido durante o presente estudo de manipulação genética envolvendo a levedura industrial JAY270.

Inativação de genes através de mutações pontuais e interrupção gênica por meio de sequências de DNA carregando consigo marcas exógenas - que por sinal são comumente encontradas em leveduras *S. cerevisiae* de laboratório, como a S288c - também não se mostravam favoráveis aos propósitos da presente invenção.

Quanto às mutações pontuais, além da possibilidade de reversão reconstituindo o gene selvagem, a sequência do gene de interesse não é retirada do genoma da levedura, o que poderia se tornar um empecilho na adoção de estratégias para o futuro melhoramento genético da linhagem JAY270. Vale lembrar que antes de retornar à usina, a levedura melhorada deverá ter a sua condição de biossíntese de uracila reconstituída pela reintrodução do gene *URA3*. Nesse momento não seria conveniente que a sequência do gene *URA3* inativado pela mutação pontual interferisse na estratégia de reintrodução do gene selvagem, até mesmo sabendo que boa parte das técnicas de manipulação genética de leveduras conta com o evento de recombinação entre sequências homólogas (Ausubel *et al.*, 1998, John Wiley & Sons Inc., intitulado “Current Protocols in Molecular Biology”). Por outro lado, a interrupção do gene *URA3* por meio de sequências de DNA contendo genes bacterianos de seleção deixaria um rastro de sequência exógena no genoma da levedura, não se mostrando atrativo perante o objetivo do invento.

Recentemente foi desenvolvido um procedimento de manipulação genética capaz de evitar muitos dos problemas mencionados acima. Conhecido como *Delitto Perfetto* (Storici *et al.*, 2001, Nature Biotechnology, 19: 773-776; Storici & Resnick, 2006, Methods Enzymol, 409: 329-45; Patente: Derwent n. WO 2003012036-A2), ele permite deletar e/ou modificar genes com a utilização de apenas um único marcador genético para um número indefinido de manipulações, evitando a criação de múltiplas mutações como tradicionalmente feito com as linhagens *S. cerevisiae* de laboratório.

Com esse sistema é possível produzir uma marca de auxotrofia em uma linhagem selvagem atuando diretamente no seu cromossomo, de maneira bem mais específica do que as técnicas de mutagênese que têm sido aplicadas em linhagens de laboratório como a S288c. Por estas razões, o *Delitto Perfetto* é um procedimento de manipulação considerado “minimamente invasivo”, pois além de realizar a modificação genética de interesse sem deixar marcas de transgenes ao final do processo no genoma da levedura, ele permite também que a linhagem manipulada tenha as suas características naturais preservadas.

O sistema *Delitto Perfetto* se procede por meio de duas etapas (Storici *et al.*, 2001, *Nature Biotechnology*, 19: 773-776; Storici & Resnick, 2006, *Methods Enzymol*, 409: 329-45).

Na primeira etapa, um cassete CORE contendo os genes *klURA3* da levedura *Kluyveromyces lactis* (contra-seleção) e *KanMX4* (marca de seleção para geneticina) é inserido no *locus* a ser deletado.

A segunda etapa consiste na retirada do cassete CORE utilizando o fragmento homólogo de reparo, que é composto apenas pelas regiões flangeadoras do *locus* de interesse, não havendo quaisquer marcas genéticas de seleção. Assim, a contra-seleção em 5-FOA para a perda da marca *klURA3*, seguida pelo teste para a perda da marca *KanMX4*, indica a deleção da região de interesse, diminuindo a quantidade de falsos positivos.

Nesse sentido, a técnica do *Delitto Perfetto* foi incorporada na realização da presente invenção, principalmente por permitir a retirada das marcas genéticas de seleção usadas durante o processo de manipulação da linhagem JAY270.

Métodos de manipulação genética baseados na engenharia genética são bastante específicos, atuando de maneira mais precisa no gene alvo, ao contrário das técnicas mencionadas anteriormente.

A engenharia genética baseada no *Delitto Perfetto* mostrou ser um procedimento de manipulação genética ideal para a concretização da presente invenção, já que ele poderia auxiliar o trabalho de modificação genética na busca da auxotrofia para uracila, preservando a complexa arquitetura genômica da linhagem JAY270.

O fato de se conseguir preservar as características naturais da linhagem industrial, como boa capacidade fermentativa e ótima adaptação ao ambiente estressante das dornas, será determinante para a viabilidade dos estudos a serem realizados visando o melhoramento genético da linhagem, aumentando, assim, as chances de sucesso da linhagem modificada em futuras aplicações industriais (Argueso & Pereira, 2010, *International Sugar Journal*, intitulado “Perspective: Indigenous sugarcane yeast strains as ideal biological platforms for the delivery of next generation biorefining Technologies”, 112: 86-89).

O melhoramento genético de leveduras tem sido satisfatoriamente alcançado em inúmeros trabalhos (Cunha *et al.*, 2006, *Yeast Res.*, 6(2): 280-7; Wisselik *et al.*, 2007, *Appl. Envir. Microbiol.*, 73: 4881-4891), com linhagens de laboratório.

Todavia, o mesmo não ocorre com as linhagens industriais, que são as leveduras realmente empregadas nos processos de fermentação em larga escala, visto que são de difícil “domesticação” ou não apresentam uma “fácil manipulação genética”.

O fato dessas linhagens serem de difícil manipulação genética acaba sendo uma barreira para o melhoramento de suas características de interesse fermentativo-industrial.

Sumário da Invenção

Para solucionar os problemas acima mencionados, a presente invenção propiciará vantagens significativas com o desenvolvimento de uma levedura mutante nulo *ura3* da cepa industrial pedra-2 (*Saccharomyces cerevisiae*), gerada por engenharia genética, para uso industrial pela introdução de genes de interesse possibilitando um aumento do seu desempenho e apresentando uma relação custo/benefício mais favorável.

A presente invenção se refere a uma linhagem de levedura selvagem *Saccharomyces cerevisiae* auxotrófica para uracila, resultante da deleção completa das duas cópias do gene *URA3* do seu genoma diplóide.

A linhagem modificada foi denominada de FGY050. A linhagem modificada JAY270, que é um isolado da Pedra-2, principal cepa industrial brasileira de fermentação alcoólica em larga escala, sendo reconhecida pela sua extraordinária capacidade de produção de álcool a partir da cana-de-açúcar sob as condições de alto estresse das dornas.

A levedura mutante auxotrófica para uracila da presente invenção foi gerada através de uma estratégia de engenharia genética bastante específica, que permitiu somente a deleção completa do gene *URA3* (incluindo promotor e região codante), sem qualquer modificação adicional e sem deixar quaisquer marcas de transgenia, tais como os genes marcadores bacterianos de resistência a antibióticos utilizados durante o processo de manipulação genética. Portanto, a linhagem auxotrófica, aqui gerada, teve as suas características de interesse fermentativo-industrial preservadas, sendo idêntica à sua cepa original Pedra-2, exceto pela deleção da sequência do gene *URA3*.

Descrição Detalhada da Invenção

Embora a presente invenção possa ser suscetível a diferentes modalidades, é mostrada nos desenhos e na seguinte discussão detalhada, uma modalidade preferida com o entendimento de que a presente divulgação deve ser considerada uma exemplificação dos princípios da invenção e não pretende limitar a presente invenção ao que foi ilustrado e descrito aqui.

Tendo em vista as considerações elucidadas acima, é mais conveniente e, apropriado, segundo os objetivos da presente invenção, desenvolver uma linhagem de levedura industrial apresentando auxotrofia para um gene marcador. A linhagem JAY270 foi precisamente escolhida por se tratar de um isolado da Pedra-2, que é a cepa de fermentação em larga escala

mais importante do país e, por possuir grande potencial para o melhoramento genético de suas características fermentativas.

Com o auxílio do *Delitto Perfetto*, apenas linhagens auxotróficas para uracila podem ser geneticamente melhoradas. As leveduras que foram modificadas geneticamente no próprio artigo do *Delitto Perfetto* eram todas auxotróficas para uracila. Como a retirada do cassete CORE (*kanMX4* e κ *URA3*) pelo fragmento homólogo de reparo é verificada pela perda da condição URA⁺ e, assim pela perda da capacidade de reconhecer o composto 5-FOA, não pode haver nenhuma cópia excedente do gene URA3 no genoma da levedura, que não somente a do próprio cassete CORE. Se a levedura for prototrófica não é possível aplicar a técnica do *Delitto Perfetto*, ao menos que o gene alvo seja o próprio gene URA3, como é o caso deste presente invento.

Se a levedura for haplóide é possível deletar o gene URA3 e retirar todas as marcas de seleção apenas com as duas etapas de manipulação genética da estratégia do *Delitto Perfetto*. No entanto, se a levedura for diplóide ou poliplóide é preciso elaborar uma nova estratégia em cima da estratégia do *Delitto Perfetto* para se conseguir a deleção completa das cópias do gene URA3 sem deixar nenhuma marca genética de seleção ao final do processo.

No caso de uma linhagem HAPLÓIDE URA⁺, a estratégia para a deleção completa da sua única cópia do gene URA3 seria: 1º) Deleção do gene *URA3* com o cassete CORE (*kanMX4* e κ *URA3*), onde a seleção é feita em meio contendo geneticina; neste ponto, a linhagem ainda é URA⁺, devido a presença do κ *URA3* dentro do cassete e 2º) Remoção do cassete CORE com o fragmento homólogo de reparo, cuja seleção é feita em meio contendo 5-FOA, já que a levedura transformante deverá ter perdido o gene marcador κ *URA3* e, assim, a condição URA⁺ e a capacidade de reconhecer o 5-FOA, sendo incapaz de produzir o composto tóxico).

No caso de uma levedura DIPLÓIDE URA⁺, como é o caso da JAY270 - alvo de manipulação da presente invenção - não é possível seguir totalmente o esquema da técnica *Delitto Perfetto*, tal como exemplificado para uma levedura haplóide, conforme acima referido.

No caso da JAY270, alvo de manipulação genética da presente invenção, que é uma levedura diplóide, foi preciso considerar as duas cópias do gene URA3. Se tratando de uma levedura diplóide é preciso se referir como 1ª e 2ª cópia do gene *URA3*. Extrapolando, seria como se fosse o genoma de duas leveduras haplóides dentro de uma única célula. A manipulação genética de cada cópia é independente, mas ambas precisam ser consideradas dentro da mesma estratégia.

Para entender melhor a diferença na estratégia entre uma levedura haplóide e diplóide é válido a seguinte suposição: uma das duas cópias do gene *URA3* de uma levedura diplóide, como a JAY270, sendo removida do genoma da linhagem pelo cassete CORE (*kanMX4* e *κ_IURA3*). A seleção da levedura transformante deve ser feita em meio contendo geneticina dada a presença do gene *kanMX4* dentro do cassete CORE. Neste ponto, a levedura transformante ainda é *URA+*, só que neste caso (diferentemente do caso acima da levedura haplóide), por dois motivos diferentes. A levedura transformante ainda será *URA+* pois ela ainda terá uma cópia do gene *URA3* ativa no seu genoma (2º *locus*), além do gene *κ_IURA3* do próprio cassete de integração (CORE). A presença desta 2ª cópia do gene *URA3* no genoma da levedura diplóide faz muita diferença na elaboração da estratégia de manipulação genética para a deleção dos seus genes.

Assim, no caso acima referido, não seria possível efetuar a retirada do cassete CORE com o fragmento homólogo de reparo, já que a 2ª cópia intacta do gene *URA3* no seu genoma impediria a contra-seleção em 5-FOA. Isto, pois mesmo perdendo a marca *κ_IURA3*, a levedura seria capaz de reconhecer o 5-FOA, devido a permanência do gene *URA3* no 2º *locus*, produzindo o composto tóxico e morrendo em seguida. Assim sendo, não seria possível dar continuidade ao processo de deleção das duas cópias do gene *URA3* do genoma de uma levedura diplóide com o auxílio apenas do Delitto Perfetto.

Foi exatamente por essa dificuldade que foi montado uma estratégia diferente para a remoção das duas cópias do gene *URA3* da linhagem JAY270.

A primeira etapa da domesticação da JAY270 é fundamental para a viabilidade do processo. Esta primeira etapa é importante por permitir remover uma das cópias do gene *URA3* e, ao mesmo tempo, por não introduzir nenhum gene *URA3* no *locus*. Ela é importante principalmente pensando nas etapas seguintes, ou seja, visando a continuidade do processo.

De acordo com a presente invenção, uma linhagem de levedura selvagem *Saccharomyces cerevisiae* auxotrófica para uracila, resultante da deleção completa das duas cópias do gene *URA3* do seu genoma diplóide é revelado.

A linhagem auxotrófica gerada é incapaz de crescer em meio sem uracila devido à deleção das duas cópias do gene *URA3*. Logo, a linhagem gerada pela presente invenção é passível de ser geneticamente manipulada pela introdução de genes diversos acompanhados pelo gene marcador *URA3*, já que os transformantes modificados podem ser reconhecidos pela sua capacidade de crescimento em meio de cultura sem uracila.

Estas cepas assim modificadas podem, por sua vez, ter o gene marcador *URA3*

novamente removido a partir da seleção em meio de cultura contendo a droga 5-FOA, voltando a ser uma cepa mutante auxotrófica para uracila (genótipo *ura3*) e, portanto, receptiva a novos *rounds* de transformação. A linhagem auxotrófica gerada neste invento permite assim a utilização recorrente do gene marcador *URA3*, por um número indefinido de vezes, de modo que é possível introduzir genes específicos dentro do seu genoma, sendo possível construir até mesmo novas vias metabólicas, capazes de produzir diversos compostos de interesse comercial.

Dessa forma, a presente invenção apresenta não só grande potencial na produção de maiores quantidades de álcool, como também pode ser utilizada para a produção de outros compostos de interesse industriais, sejam eles orgânicos ou não, protéicos ou não, uma vez que a linhagem modificada apresenta grande versatilidade.

Nesse contexto, é proposto pela presente invenção “domesticar” uma levedura selvagem de alta importância para as indústrias, neste caso a JAY270, que é um isolado da cepa Pedra-2, para facilitar estudos que busquem o melhoramento genético da linhagem visando maior produção de etanol combustível e também a confecção de outros produtos de interesse industrial. Em outras palavras, o presente invento teve como objetivo desenvolver uma levedura selvagem de fácil manipulação genética, plataforma biológica versátil para a introdução de características fermentativas desejáveis para o seu emprego nas indústrias.

A Pedra-2 é capaz de produzir álcool a um baixo custo, fator fundamental para a obtenção de commodities, comercialmente competitivos, pela via de fermentação.

A presente invenção desenvolveu uma levedura de fácil manipulação genética com grande potencial fermentativo para aplicações industriais.

Por esta razão, optou-se por trabalhar com a linhagem JAY270, que é um isolado da PE-2, cepa brasileira de fermentação em larga escala, reconhecida pela sua extraordinária capacidade de produção de álcool a partir da cana-de-açúcar sob condições de alto estresse das dornas.

Métodos de manipulação genética foram precisamente escolhidos para gerar a auxotrofia para uracila sem interferir no genoma da levedura, garantindo, assim, que suas características fermentativas naturais fossem preservadas.

Foi desenvolvida pela presente invenção uma linhagem de levedura selvagem da espécie *Saccharomyces cerevisiae* com auxotrofia para uracila, após deleção completa das duas cópias do gene *URA3* (cromossomo V) do genoma diplóide da levedura JAY270, através de métodos específicos de engenharia genética.

A levedura manipulada é geneticamente igual à sua parental selvagem JAY270, exceto pela auxotrofia gerada que resultou na sua incapacidade de crescer em meio sem uracila. Isto, pois a enzima codificada pelo gene *URA3* está envolvida na via de biossíntese de uracila, de modo que a deleção das duas cópias gênicas causa a interrupção desta via metabólica.

Segue abaixo a sequência de nucleotídeos codante do gene *URA3* com a identificação do número de SEQ ID. Esta sequência está disponível no banco de dados da levedura *Saccharomyces cerevisiae*, conhecido como SGD (<http://www.yeastgenome.org/>).

SEQ ID 01: YEL021W

```
ATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCAC
GAAAAGCAAACAACTTGTGTGCTTCATTGGATGTTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGT
CCCCAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGAGGGGCACAGTTAAGCCGCTAAAG
GCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAG
TACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGGTGGGCCCAGGTATTGTT
AGCGGTTTGAAGCAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATGCAAG
GGCTCCCTAGCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATCGGCCTTT
ATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGAC
AAGGGAGACGCATTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGA
AGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTG
AGAAGATGCGGCCAGCAAAACTAA
```

Leveduras selvagens são naturalmente prototróficas, ou seja, não apresentam marcas de auxotrofia. Por esse motivo a manipulação genética de leveduras selvagens acaba requerendo a utilização de estratégias que contem com o emprego de marcas genéticas bacterianas de resistência aos antibióticos para a seleção dos transformantes que carregam consigo a modificação de interesse.

Se tratando de uma levedura diplóide, a estratégia de manipulação genética envolvendo a linhagem JAY270 contou com a utilização de dois genes bacterianos de resistência a drogas. Além do gene *kanMX4* (resistência à geneticina, G418) que está presente no plasmídeo pCORE do sistema *Delitto Perfetto* (Storici *et al.*, 2001; Storici & Resnick, 2006) foi necessário o uso de uma segunda marca, que neste caso foi o gene *hphMX4* (resistência à higromicina) presente no plasmídeo pEAI, que é semelhante ao pAG32 (Goldstein & McCusker, 1999, *Yeast*, 15: 1541-1553). Embora o gene marcador *hphMX4* tenha sido escolhido, poderiam se utilizados outros genes marcadores de resistência a outros tipos de drogas (Kanamicina, por

exemplo).

Cinco etapas de manipulação genética foram necessárias para que a linhagem JAY270 passasse a apresentar auxotrofia para uracila, sem possuir no final do processo quaisquer marcas genéticas de transgenia no seu genoma conforme apresentado na Figura 1.

As cópias do gene *URA3* foram deletadas com cassetes de integração que carregavam consigo marcas genéticas de seleção diferentes (*hphMX4* ou *kanMX4*). Como a levedura possui a capacidade de reconhecer sequências de DNA que apresentam homologia ao seu genoma, por recombinação é possível ocorrer integração de cassetes integrativos com conseqüente deleção de cópias gênicas.

Além das etapas de deleção das duas cópias do *URA3* foram necessárias outras etapas até que as marcas *hphMX4* e *kanMX4* fossem retiradas do genoma da levedura JAY270. Esse cuidado foi super importante para que a levedura final modificada não apresentasse marcas de transgenia e, portanto, não fosse considerada como um microrganismo transgênico, ainda mais sabendo que em um futuro próximo ela será utilizada em estudos de melhoramento genético para ser empregada em aplicações industriais, com possíveis fins comerciais. Assim, nesse futuro próximo, não haverá a necessidade de uma usina ou empresa gastar com controles de disseminação de OGM transgênicos no meio ambiente, já que estará usando um organismo não transgênico.

A primeira etapa (Etapa 01) do processo de domesticação da JAY270 apresentado na Figura 1 consistiu na deleção completa de uma das duas cópias do gene *URA3* da levedura, utilizando um cassete de integração, denominado “cassete Hph”, o qual possuía a marca genética *hphMX4* de resistência à droga higromicina.

A linhagem transformante foi denominada como FGY010 e seu genótipo é igual ao da JAY270 exceto pela deleção de uma cópia do gene *URA3* (JAY270 *ura3Δ::Hph/URA3*).

Na Etapa 02, a segunda cópia do gene *URA3* da levedura FGY010 foi deletada com o auxílio de um cassete de integração, denominado “cassete CORE”, o qual possuía o gene *kanMX4* de resistência à droga geneticina (G418) e, também, o gene *κ_IURA3* da levedura *Kluyveromyces lactis*. A linhagem transformante foi denominada como FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*).

A terceira etapa (Etapa 03) consistiu na retirada do cassete CORE do genoma da levedura FGY020 utilizando o fragmento homólogo de reparo, que é composto apenas pelas sequências das regiões upstream e downstream do gene *URA3*, não apresentando nenhuma marca de seleção no seu interior. O transformante foi denominado de FGY030 (JAY270

ura3Δ::Hph/ura3Δ). Como este fragmento não possui nenhuma marca de seleção, há a necessidade da presença do gene *κ_IURA3* dentro do cassete CORE para que a levedura FGY030 possa ser selecionada em meio contendo 5-FOA, no qual crescem somente leveduras incapazes de sintetizar uracila e, assim, impossibilitadas de metabolizar o 5-FOA e produzir o composto tóxico.

A linhagem FGY030 já apresenta auxotrofia para uracila, porém ela possui uma marca de transgenia (*hphMX4*) indesejável no seu genoma. Assim, as duas etapas finais (Etapa 04 e 05) foram realizadas para retirar a marca *hphMX4* do genoma da levedura FGY030.

Na Etapa 04, a marca genética *hphMX4* foi retirada do genoma da levedura FGY030 através da integração do cassete CORE (FGY040 = JAY270 *ura3Δ::CORE/ura3Δ*). Ao transformar a FGY030 com o cassete CORE a levedura passou a ser capaz de crescer em meio sem uracila, dada a presença do *κ_IURA3*. A retomada da condição prototrófica (URA3+) foi fundamental para que a quinta e última etapa pudesse ser realizada, já que a seleção seria feita em meio contendo 5-FOA para a perda do *κ_IURA3*.

Assim, a última etapa do processo de manipulação da JAY270 (Etapa 05) resumiu-se à retirada do cassete CORE do genoma da levedura FGY040 e, assim, das marcas genéticas *kanMX4* e *κ_IURA3*, utilizando novamente o fragmento homólogo de reparo (FGY050 = JAY270 *ura3Δ/ura3Δ*).

Antes de iniciar o processo de domesticação da JAY270 foi construído dois cassetes de integração (Hph e CORE) e, também, o fragmento homólogo de reparo.

O cassete Hph foi construído a partir do plasmídeo pEI, que é semelhante ao plasmídeo pAG32 descrito por Goldstein & McCusker, 1999, Yeast, intitulado “Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*”, 15: 1541-1553, incorporado aqui como referência em sua totalidade.

O cassete CORE foi construído a partir do plasmídeo pCORE (Storici *et al.*, 2001, Nature Biotechnology, 19: 773-776; Storici & Resnick, 2006, Methods Enzymol, 409: 329-45) incorporado aqui como referência em sua totalidade.

Construção dos cassetes integrativos Hph e CORE

Para a construção dos cassetes, foram desenhados *primers* (JAO286, JAO287 e JAO288; SEQ ID 02, 03 e 04, respectivamente) compostos por duas partes com funções distintas, sendo a terminação 3' (não sublinhada) complementar ao molde a ser amplificado nos plasmídeos e, a terminação 5' (sublinhada) complementar às regiões *upstream* e *downstream*

do *URA3* do cromossomo V da levedura (Tabela 01).

Os *primers* JAO286 e JAO287 foram utilizados na construção do cassete Hph, enquanto que os *primers* JAO286 e JAO288 foram usados na amplificação do cassete CORE, conforme mostrado na Tabela 1.

Tabela 1 - *Primers* (JAO286; JAO287 e JAO288) para a construção dos cassetes de integração Hph e CORE. (R) – primer reverse; (F) – primer forward.

Primer	Sequência (5' → 3')
JAO286 (R)	ACCCAAGTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAATCGAGC TCGTTTTCGACACTGG
JAO287 (F)	AATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTGACGT GCGCAGCTCAGGG
JAO288 (F)	AATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTCTTAC CATTAAGTTGATC

Após a amplificação por PCR os cassetes integrativos foram delimitados, de modo que as sequências dos *primers* com homologia às regiões adjacentes ao gene *URA3* passaram a compor as suas extremidades (lados direito e esquerdo), enquanto que as marcas genéticas de seleção (*hphMX4*; *kanMX4* e *κ_IURA3*) permaneceram mais ao centro conforme apresentado na Figura 02.

Portanto, ambos os cassetes possuem nas suas duas extremidades aproximadamente 45 bases contendo homologia com as regiões *upstream* e *downstream* do gene *URA3*, as quais foram fundamentais para que houvesse a integração dos cassetes no genoma da levedura selvagem JAY270 durante as transformações, por recombinação homóloga.

A construção dos cassetes integrativos Hph e CORE via reação de PCR foi feita segundo as seguintes condições: 1µl de DNA (plasmídeo), 1µl de tampão (10X), 0,3µl de MgCl₂ (50mM), 0,5µl de cada *primer* (15pmol), 2µl de dNTPs (1,25mM) e 0,1µl de taq DNA polimerase de alta fidelidade (50U/µl) (*Invitrogen*) para um volume final de 10µl.

O programa utilizado foi: um ciclo inicial de desnaturação a 94°C por 5 minutos, seguido por 35 ciclos de 94°C por 30 segundos, temperatura de anelamento por 30 segundos (57 a 68°C para os *primers* JAO286/JAO287 e 51 a 63°C para os *primers* JAO286/JAO288) e 68°C por 1 minuto e 30 segundos (cassete Hph) ou por 3 minutos (cassete CORE), além da extensão final de 68°C por 5 minutos.

O fragmento gerado pela construção do cassete Hph com os *primers* JAO286/JAO287 é

de 1678pb (resultado não mostrado). Por outro lado, um fragmento de 3285pb foi gerado na construção do cassete CORE com os *primers* JAO286/JAO288 (resultado não mostrado).

Estas reações seguiram o modelo padrão de amplificação segundo as condições programadas no equipamento de rotina no laboratório, denominado PCR. Então, por exemplo, 94 °C é a temperatura para que haja a abertura da dupla fita de DNA (que é fixa) que permite em seguida que os *primers* se anelem. A temperatura de anelamento, por sua vez, é a única temperatura que vai de acordo com a sequência do *primer* utilizado.

Construção do Fragmento Homólogo

A construção do fragmento homólogo de reparo se deu através de quatro passos como mostrado na Figura 03, onde do lado esquerdo estão os quatro passos necessários para a sua construção e do lado direito estão os produtos de PCR correspondentes. Os pares JAO333 e JAO334 se anelam às regiões *upstream* e *downstream* do gene *URA3* do genoma da JAY270.

Inicialmente (Passo I), os *primers* JAO333/JAO334 foram empregados na amplificação de uma sequência de DNA, a partir do genoma da levedura selvagem JAY270. Este fragmento de DNA de 1766pb continha a sequência do gene *URA3* e 300 a 600 bases das suas regiões *upstream* e *downstream*. É importante lembrar que a parte de interesse está justamente nas regiões flangeadoras do gene *URA3*.

A reação de PCR foi feita segundo as seguintes condições: 1µl de DNA genômico da JAY270, 1µl de tampão (10X), 0,3µl de MgCl₂ (50mM), 0,14µl de cada um dos *primers* (15pmol), 2µl de dNTPs (1,25mM) e 0,1µl de taq DNA polimerase de alta fidelidade (50U/µl) (*Invitrogen*) para um volume final de 10µl. O programa utilizado foi: um ciclo inicial de desnaturação a 94°C por 2 minutos, seguido por 35 ciclos de 94°C por 30 segundos, 47 a 56°C por 30 segundos, 68°C por 2 minutos e uma extensão final de 68°C por 10 minutos.

Feito isso, esta sequência de DNA, de 1766pb, foi digerida (Passo II) pelas enzimas de restrição PstI e NsiI, gerando três fragmentos (311pb; 610pb e 825pb). Os dois fragmentos menores (311pb e 610pb, que representam as regiões *upstream* e *downstream* do gene *URA3*) foram isolados do gel de agarose, purificados (Kit *Qiagen*) e ligados (T4 DNA ligase) (Passo III). Das três combinações possíveis (311pb + 311pb; 610pb + 311pb e 610pb + 610pb), apenas a de interesse (610pb + 311pb) foi isolada do gel de agarose e purificada. Este fragmento de 921pb foi utilizado em novas reações de PCR (JAO333/JAO334) para aumentar a quantidade do seu produto a ser utilizado nas suas etapas de transformação (Passo IV).

Vale lembrar, que este fragmento de 921pb já é o fragmento homólogo de reparo de

interesse. A composição dele nada mais é do que o fragmento inicial de 1766pb sem a sequência do gene *URA3*. Assim, o fragmento homólogo de reparo é constituído apenas pelas regiões “externas” do gene (regiões flangeadoras), não havendo nenhuma marca no seu interior. É exatamente por isso que ele foi usado na retirada das marcas genéticas de seleção (*hphMX4*, *kanMX4* e *κ_IURA3*) usadas durante o processo de domesticação da JAY270.

Para confirmar as transformações relativas a cada etapa do processo de domesticação da JAY270 foram desenhados diversos *primers* conforme mostrado na Tabela 02. Pares de *primers* (JAO259/JAO260; JAO333/JAO334 e JAO332/JAO336) localizados nas regiões *upstream* e *downstream* do gene *URA3* do genoma da levedura *S. cerevisiae* apresentados na Figura 04 permitem amplificar o gene selvagem, bem como verificar a integração correta dos cassetes após as transformações. Foram desenhados também *primers* internos (JAO14 e JAO15) aos cassetes que combinados com *primers* externos (ao *locus URA3*) de sentido contrário, permitem verificar se a integração ocorreu no local desejado. Pode-se observar que o *primer* JAO14 é exclusivo do cassete CORE, enquanto que o *primer* JAO15 é interno a ambos os cassetes (Hph e CORE).

Assim, diversas combinações de *primers* permitiram constatar deleções das cópias do *URA3*, integração dos cassetes Hph e CORE, bem como retirada das marcas genéticas de seleção pela ação do fragmento homólogo de reparo.

Tabela 02 - *Primers* para confirmar as transformações. Os *primers* (JAO259; JAO260; JAO333; JAO334; JAO332 e JAO336) estão localizados nas regiões *upstream* e *downstream* do gene *URA3* do genoma da levedura *S. cerevisiae* e os *primers* (JAO14 e JAO15) são internos aos cassetes de integração

<i>Primer</i>	SEQ ID	Sequência (5' → 3')
JAO259	05	ATGTGGTGTTGAAGAAACATG
JAO260	06	AATCATTACGACCGAGATTCC
JAO333	07	GGGAAGACAAGCAACGAAACG
JAO334	08	ACCAGATTAGAGTACAAACGC
JAO332	09	TACCGTTGCTAAATTCTGAGTG
JAO336	10	TGATGTTGTGAAGTCATTGAC
JAO14	11	AGGAGGGTATTCTGGGCCTCCATG
JAO15	12	ATGCGAAGTTAAGTGCGCAGAAAG

Inúmeras tentativas de transformação foram realizadas, utilizando-se o cassete Hph construído a partir do plasmídeo pEAI, sem que houvesse deleção do gene de interesse. Isto pode estar relacionado com o número reduzido de bases homólogas existente entre o pequeno fragmento, presente nas extremidades do cassete integrativo e as regiões que flanqueiam o gene *URA3* no genoma da levedura.

A estratégia usada na construção de ambos os cassetes integrativos restringiu a quantidade de bases homólogas às regiões adjacentes ao gene *URA3*, sendo que neste caso limitou-se a aproximadamente 45 bases de cada lado dos cassetes.

Tendo em vista que a eficiência da transformação poderia estar diretamente relacionada com o número de bases homólogas presente nas extremidades do cassete integrativo, que participam do evento de recombinação homóloga, foi efetuado um procedimento alternativo para solucionar tal limitação e facilitar o processo de manipulação genética da JAY270.

Este procedimento consistiu em:

(1) transformar a linhagem padrão de laboratório S288c, que é de fácil manipulação genética, com o mesmo cassete integrativo construído a partir do plasmídeo e,

(2) construir um novo cassete integrativo, com número de bases homólogas aumentado, tendo como base o genoma da levedura S288c transformada, utilizando pares de *primers* mais externos (JAO333 e JAO334) ao gene *URA3* como mostrado na Figura 4, onde os cassetes Hph e CORE estão integrados no *locus URA3* do genoma da levedura JAY270, representado por linhas pontilhadas. Os *primers* JAO259, JAO260, JAO333, JAO334, JAO332 e JAO336 estão localizados nas regiões upstream e downstream do gene *URA3* enquanto que os *primers* JAO14 e JAO14 são internos aos cassetes.

Assim, a levedura de laboratório S288c foi inicialmente transformada com o cassete Hph construído a partir do plasmídeo pEAI e, em seguida, o seu genoma modificado serviu de base para a construção de um novo cassete Hph, contendo, agora, nas suas extremidades, um fragmento bem maior de homologia às regiões adjacentes ao gene *URA3*. Com esse procedimento, este novo cassete Hph passou a ter nas suas duas extremidades um fragmento de 300 bases de homologia às regiões adjacentes ao gene *URA3* (resultados não mostrados).

A construção deste novo cassete Hph, contendo um fragmento bem maior de homologia às regiões adjacentes ao gene *URA3*, permitiu retomar a Etapa 01 da domesticação da linhagem JAY270.

Embora a construção deste cassete integrativo tenha exigido maior dedicação e tempo, este procedimento mostrou-se bastante eficiente, já que a transformação desejada foi

alcançada com sucesso.

Os transformantes resistentes à droga higromicina e que, portanto, possuíam o cassete Hph integrado no seu genoma, foram testados a fim de verificar se a integração havia ocorrido no local desejado, tendo por consequência, deletado uma das cópias do gene *URA3* do genoma da JAY270.

A transformação desejada foi confirmada por PCR, utilizando os *primers* JAO259, JAO260 e JAO15. De 10 candidatos testados, sete eram os transformantes FGY010 (JAY270 *ura3Δ::Hph/URA3*).

Pode-se observar a amplificação das duas bandas esperadas para os transformantes de interesse conforme mostrado na Figura 05, onde M é o marcador molecular (100pb); os números 1 e 2 são os dois transformantes de interesse (FGY010). Uma delas (1000pb) corresponde a uma das cópias do gene *URA3* intacto, amplificada pelos *primers* JAO259/JAO260. A outra (200pb), gerada pelos *primers* JAO259/JAO15, corresponde à banda esperada pela integração do cassete Hph e conseqüente deleção de uma das cópias do gene. Como a reação de PCR foi realizada usando três *primers* em conjunto, há uma competição natural entre eles, de modo que a banda de menor tamanho (200pb) tende a mostrar-se mais intensa.

A confirmação da Etapa 01 foi feita segundo as seguintes condições: 1μl de DNA, 1μl de tampão (10X), 0,3μl de MgCl₂ (50mM), 0,14μl de cada um dos *primers* (15pmol), 2μl de dNTPs (1,25mM) e 0,1μl de taq DNA polimerase (50U/μl) (*Invitrogen*) para um volume final de 10μl. O programa utilizado foi: um ciclo inicial de desnaturação a 94°C por 2 minutos, seguido por 35 ciclos de 94°C por 30 segundos, 49 a 62°C por 30 segundos e 72°C por 1 minuto e 30 segundos, além da extensão final de 72°C por 10 minutos.

Os produtos de PCR foram submetidos à eletroforese em gel de agarose (1 a 1,5%) por um período de 1 a 2 horas, utilizando-se corrente entre 60 a 100V e, visualizados em fotodocumentador após marcação com brometo de etídeo.

O resultado obtido via PCR foi corroborado ao verificar a segregação mendeliana das marcas genéticas *URA3* e *ura3Δ::Hph* nos quatro esporos da linhagem transformante FGY010 (JAY270 *ura3Δ::Hph/URA3*). Para isso, a linhagem FGY010 foi plaqueada em meio de esporulação e, em seguida, submetida à dissecação de tétrades através de um sistema de micromanipulação acoplado a um microscópio (*Nikon Eclipse E400*). Quando plaqueados em meio com deficiência em uracila, apenas dois dos quatro esporos da levedura transformada foram capazes de crescer, indicando a presença do gene *URA3*. Por outro lado, apenas os

esporos incapazes de crescer neste meio mostravam crescimento normal em meio contendo higromicina, evidenciando a presença da marca *ura3Δ::Hph*.

Estes resultados indicam a deleção de apenas uma cópia do gene *URA3* da linhagem JAY270, conforme o esperado, confirmando o término da Etapa 01 do processo de domesticação.

Com a finalização desta primeira etapa, foi possível iniciar a Etapa 02 do processo de manipulação genética da JAY270, que consistiu na deleção da segunda cópia do gene *URA3* do genoma da levedura FGY010 (JAY270 *ura3Δ::Hph/URA3*) utilizando o cassete CORE conforme mostrado na Figura 1.

Para evitar as dificuldades encontradas na primeira etapa do processo, a construção do cassete CORE seguiu o mesmo procedimento utilizado na obtenção do cassete Hph contendo nas suas extremidades um fragmento bem maior de homologia às regiões adjacentes ao *URA3* (resultado não mostrado).

Após a sua construção, o cassete CORE de maior fragmento de homologia (~300 bases) permitiu a realização bem sucedida da Etapa 02 da domesticação da JAY270.

Feita a segunda seleção em meio suplementado com geneticina (G418), os transformantes candidatos foram testados a fim de verificar a integração correta do cassete CORE no genoma da levedura FGY010 (JAY270 *ura3Δ::Hph/URA3*), resultando, apenas, na deleção da segunda cópia do gene *URA3*. Dos três candidatos obtidos, dois apresentaram a transformação desejada mostrado na Figura 06, onde M é o marcador molecular (1Kb *Fermentas*); os números 1, 2 e 3 indicam os transformantes candidatos. A banda de 1900pb (JAO260/JAO14) indica a integração do cassete CORE no genoma da levedura FGY010. Os *primers* JAO259/JAO260 mostram a permanência da marca *hphMX4* (1800pb) em apenas dois candidatos (1 e 2), enquanto que o outro candidato (3) manteve indesejadamente a segunda cópia do gene *URA3* intacta (1000pb).

Foram realizadas reações de PCR com pares de *primers* diferentes para constatar a integração do cassete CORE (JAO260/JAO14), e também, para verificar a permanência da marca genética *hphMX4* (JAO259/JAO260) no genoma da levedura FGY010. A integração do cassete CORE no genoma dos três candidatos foi confirmada através da amplificação da banda de 1900pb, utilizando-se os *primers* JAO260/JAO14. No entanto, a reação de PCR com os *primers* JAO259/JAO260 mostrou que a marca genética *hphMX4* (1800pb) estava presente em apenas dois deles (candidatos 1 e 2). Este mesmo par de *primers* mostrou que o outro candidato (3) manteve a segunda cópia do gene *URA3* intacta, de modo que a integração do

cassete CORE conduziu a retirada indesejada da marca genética *hphMX4*. Isto explica o fato dos dois primeiros candidatos (1 e 2) terem se mostrado resistentes à higromicina, enquanto o último (3) apresentou sensibilidade a esta droga.

As seguintes condições de PCR foram necessárias para confirmar a integração do cassete CORE com o auxílio dos *primers* JAO260 e JAO14: 1µl de DNA, 1µl de tampão (10X), 0,3µl de MgCl₂ (50mM), 0,14µl de cada um dos primers (15pmol), 2µl de dNTPs (1,25mM) e 0,1µl de Taq DNA polimerase (50U/µl) (*Invitrogen*) para um volume final de 10µl. O programa utilizado foi: um ciclo inicial de desnaturação a 94°C por 2 minutos, seguido por 35 ciclos de 94°C por 30 segundos, 54 a 63°C por 30 segundos e 72°C por 2 minutos, além da extensão final de 72°C por 10 minutos.

A reação de PCR utilizando os *primers* JAO259/JAO260 seguiu as mesmas condições apresentadas na confirmação da Etapa 01 do processo de domesticação. Os produtos de PCR foram submetidos à eletroforese em gel de agarose (1 a 1,5%) por um período de 1 a 2 horas, utilizando-se corrente entre 60 a 100V e, visualizados em fotodocumentador após marcação com brometo de etídeo.

Além das reações de PCR, os transformantes da Etapa 02 foram submetidos a testes fenotípicos em diversos meios seletivos para verificar a presença das marcas genéticas (*hphMX4*; *kanMX4* e *κ_IURA3*) no genoma da linhagem FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*).

Dois dos quatro esporos de uma tétrade da levedura FGY020 foram capazes de crescer em meio contendo geneticina e em meio deficiente em uracila, sendo incapazes de crescer em meio contendo higromicina, evidenciando a presença do cassete CORE (*kanMX4* e *κ_IURA3*). Ao contrário, os outros dois esporos da mesma tétrade apresentaram resistência à higromicina, não crescendo em meio suplementado com geneticina e nem em meio sem uracila, indicando a presença do cassete Hph (*hphMX4*). Isso corrobora o término da etapa 02 da domesticação da JAY270.

É válido lembrar, que neste momento, a linhagem FGY020 apresenta marcas de resistência aos dois antibióticos (higromicina e geneticina), possuindo também o gene *κ_IURA3* ativo no seu genoma, o qual é fundamental para a continuidade do processo.

A Etapa 03 do processo de domesticação da JAY270 consistiu na retirada do cassete CORE da 2ª cópia do *locus URA3* do genoma da levedura FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*) com o fragmento homólogo de reparo.

Como o fragmento homólogo de reparo não possui nenhuma marca genética seletiva, a

seleção dos transformantes é verificada pela perda do gene κ *URA3* em meio contendo 5-FOA, já que somente leveduras com a via de biossíntese de uracila interrompida cresceriam nesta condição, por não metabolizarem o 5-FOA e não produzirem o composto tóxico. Portanto, a condição URA⁺ da levedura FGY020, pela presença do κ *URA3* ativo no seu genoma, foi essencial para a realização desta terceira etapa.

Esta foi à etapa de maior complexidade frente às outras já concluídas, principalmente no que diz respeito à busca e confirmação do transformante de interesse, pois ao colocar a levedura FGY020 em meio contendo 5-FOA, sem tê-la submetido a qualquer tipo de transformação, seria possível observar o crescimento de algumas colônias, em decorrência de basicamente três situações distintas que as tornariam auxotróficas para uracila e, portanto, não susceptíveis ao composto tóxico: (I) mutação natural no gene κ *URA3*; (II) inativação da transcrição do κ *URA3* e (III) recombinação mitótica, segregando parte do braço esquerdo dos cromossomos V em homozigose (*ura3Δ::Hph/ura3Δ::Hph*), não carregando consigo o *locus* contendo o cassete CORE, onde estava presente a marca do κ *URA3*.

Assim sendo, a seleção do transformante FGY030 (JAY270 *ura3Δ::Hph/ura3Δ*) deveria ser feita dentre estes três tipos de candidatos não-transformantes.

No entanto, ao fazer a seleção negativa dos candidatos em meio contendo geneticina, os mutantes naturais (I) e os decorrentes da inativação da transcrição (II) seriam facilmente eliminados, devido à presença da marca genética *kanMX4* nos seus genomas. Dessa forma, a seleção final se concentraria apenas na distinção do transformante de interesse entre os recombinantes mitóticos.

Mesmo assim, foi necessário um amplo *screening* para localizar o transformante FGY030. Dos 52 candidatos testados, 51 eram recombinantes mitóticos e apenas 01 era o transformante de interesse (JAY270 *ura3Δ::Hph/ura3Δ*).

Através de reação de PCR conforme mostrado na Figura 7, com os *primers* JAO332/JAO336, onde M é o marcador molecular (1Kb *Fermentas*); os números 1 a 11 indicam transformantes candidatos; o número 12 indica o controle positivo para a amplificação da marca *ura3Δ::Hph*, mostra que somente o candidato 6 apresentou a banda esperada para o *ura3Δ* “limpo” (1100pb). A banda de 2700pb corresponde à marca *ura3Δ::Hph*. O candidato 6 é o próprio transformante FGY030 enquanto os outros são recombinantes mitóticos.

Foi possível observar, através da Figura 7, que todos os candidatos exibiram a marca *ura3Δ::Hph*, dada a amplificação da banda de 2700pb, tal como o observado para o controle positivo (12). Entretanto, um dos candidatos (6) se destacou pela presença de uma segunda

banda, de 1100pb. Esta é justamente a banda esperada (*ura3Δ* “limpo”) para o transformante FGY030, indicando que houve a retirada do cassete CORE, e assim, das marcas *kanMX4* e *κURA3* pelo fragmento homólogo de reparo.

A diferença de intensidade entre as duas bandas do transformante FGY030 (6) é resultado natural da reação de PCR, dada a competição dos *primers* pelo produto de menor tamanho, que tende a se mostrar mais intenso.

A reação de PCR com os *primers* JAO332 e JAO336 seguiu as seguintes condições: 1μl de DNA, 1μl de tampão (10X), 0,3μl de MgCl₂ (50mM), 0,14μl de cada um dos *primers* (15pmol), 2μl de dNTPs (1,25mM) e 0,1μl de taq DNA polimerase (50U/μl) (*Invitrogen*) para um volume final de 10μl. O programa contou com um ciclo inicial de desnaturação a 94°C por 2 minutos, seguido por 35 ciclos de 94°C por 30 segundos, 47 a 56°C por 30 segundos, 72°C por 3 minutos e uma extensão final de 72°C por 10 minutos. Os produtos de PCR foram visualizados em gel de agarose (1 a 1,5%) por um período de 1 a 2 horas, utilizando-se corrente entre 60 a 100V e, visualizados em fotodocumentador após marcação com brometo de etídeo.

Para corroborar o resultado obtido nesta terceira etapa, novos testes foram realizados com alguns dos candidatos não-transformantes (1 a 5; 7 a 11 - Figura 07) e com o transformante FGY030 (6 – Figura 07), de tal forma que eles foram submetidos à esporulação e dissecação de tétrades. Os quatro esporos de diferentes tétrades dos candidatos não-transformantes apresentaram resistência quando plaqueados em meio contendo higromicina, indicando que houve segregação da marca genética *hphMX4* em homozigose nos seus parentais diplóides (JAY270 *ura3Δ::Hph/ura3Δ::Hph*). Isto está relacionado com a ocorrência de recombinação mitótica em organismos diplóides, a qual acarreta na perda de heteroziguidade.

Embora estes candidatos não tenham sido aproveitados por possuírem um padrão genotípico diferente do esperado para o transformante desejado (JAY270 *ura3Δ::Hph/ura3Δ*), foi muito importante ter constatado a ocorrência de perda de heteroziguidade nos recombinantes mitóticos, pois as características de interesse da JAY270 têm sido relacionadas à sua arquitetura genômica complexa e ao alto grau de polimorfismo em nível de nucleotídeo, de modo que qualquer perda de heteroziguidade se mostra totalmente indesejável frente aos propósitos da presente invenção.

Por outro lado, testes genotípicos e fenotípicos comprovaram a presença e ausência das marcas genéticas de seleção (*hphMX4*, *kanMX4* e *κURA3*) nos esporos da levedura transformante FGY030.

Essa preocupação da presente invenção em evitar perda de heteroziguidade e manter

ao máximo a fidelidade à cepa parental original não é encontrada em outros trabalhos. Aqui é possível observar que a presente invenção foi feita pensando na aplicação industrial futura.

O padrão genotípico dos quatro esporos de uma tétrade da linhagem FGY030 (JAY270 *ura3Δ::Hph/ura3Δ*) ficou evidente via reação de PCR com os *primers* JAO259/JAO260 conforme mostrado na Figura 08, onde M é o marcador molecular (100pb); os números 1 a 4 indicam os quatro esporos de uma tétrade da FGY030; o número 5 indica a linhagem “parental” FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*). Dois esporos (2 e 4) exibiram uma banda de 1800pb (*ura3Δ::Hph*) caracterizando a presença do cassete Hph, enquanto que os outros dois esporos (1 e 3) mostraram a amplificação da banda de 144pb (*ura3Δ* “limpo”), indicando a ausência de marcas genéticas. Aliás, as marcas genéticas (*kanMX4* e *κ_IURA3*) que foram retiradas destes esporos puderam ser visualizadas na linhagem “parental” FGY020 da etapa anterior (05), pela amplificação de uma banda de cerca de 3200pb (*ura3Δ::CORE*).

Isto comprova que o cassete CORE foi retirado da linhagem FGY020 pelo fragmento homólogo de reparo, estando assim ausente nestes dois esporos (1 e 3) da linhagem FGY030.

O padrão genotípico constatado via reação de PCR foi corroborado através de testes fenotípicos, plaqueando os esporos em diferentes meios seletivos (meio contendo higromicina; meio contendo geneticina e meio sem uracila). Dois esporos (2 e 4) (*ura3Δ::Hph*) cresceram apenas em meio contendo higromicina e, os outros dois esporos (1 e 3) (*ura3Δ* “limpo”) foram incapazes de crescer em quaisquer meios seletivos, indicando que realmente não possuíam nenhuma das três marcas de seleção (*hphMX4*, *kanMX4* ou *κ_IURA3*) (resultados não mostrados).

A confirmação do teste genotípico via reação de PCR utilizando-se os *primers* JAO259/JAO260 seguiu as mesmas condições utilizadas na confirmação da etapa 01, exceto pelo tempo de extensão que foi estendido para 3 minutos e 30 segundos, para que houvesse a amplificação de todas as bandas esperadas, inclusive à de maior tamanho, que neste caso era a banda do cassete CORE do controle positivo (5 – Figura 08).

Os produtos de PCR foram visualizados em gel de agarose (1 a 2%) por um período de 1 a 2 horas, utilizando-se corrente entre 60 a 100V e, visualizados em fotodocumentador após marcação com brometo de etídeo.

O término da etapa 03 representou um passo muito importante na domesticação da JAY270, pois neste momento, a linhagem diplóide FGY030 (JAY270 *ura3Δ::Hph/ura3Δ*) já apresentava a auxotrofia para uracila.

No entanto, o processo de manipulação não havia, ainda, chegado ao seu final, já que

ela apresentava uma marca genética de transgenia (*hphMX4*) no seu genoma, que precisava ser retirada. As duas etapas que restavam para o término do processo de domesticação, etapas 04 e 05, foram efetuadas exatamente para este propósito.

A etapa 04 consistiu na retirada do cassete Hph do genoma da levedura FGY030 (JAY270 *ura3Δ::Hph/ura3Δ*) com o auxílio do cassete CORE.

Durante a realização desta etapa, o cassete CORE poderia tanto retirar a marca genética *hphMX4* de resistência à higromicina presente na primeira cópia do *locus URA3* (*ura3Δ::Hph*) da levedura FGY030, quanto ser introduzido indesejadamente na sua segunda cópia (*ura3Δ*). Em ambos os casos a diferenciação seria somente constatada ao verificar a capacidade de crescimento dos candidatos em meio contendo higromicina, já que todos mostrariam resistência à geneticina, dada a integração do cassete CORE e, assim, da marca de seleção *kanMX4*.

Dentre os candidatos obtidos a partir da segunda seleção em meio contendo geneticina, a maior parte deles mostrou resistência também ao antibiótico higromicina, indicando que o cassete CORE havia integrado indesejadamente na segunda cópia do *locus URA3* (*ura3Δ*). Entretanto, alguns dos transformantes candidatos resistentes a G418 se mostraram sensíveis à droga higromicina e capazes de crescer em meio sem uracila, evidenciando que o cassete Hph havia sido retirado da primeira cópia do *locus URA3*, através da introdução do cassete CORE e, assim, das marcas genéticas *kanMX4* e *klURA3*.

O padrão genotípico destes últimos candidatos foi verificado através de uma reação de PCR com os *primers* JAO260 e JAO14 mostrado na Figura 09, a fim de constatar a presença do cassete CORE no genoma dos transformantes de interesse FGY040 (JAY270 *ura3Δ::CORE/ura3Δ*), onde M é o marcador molecular (ladder 100pb); o número 1 indica a linhagem “parental” FGY030 (JAY270 *ura3Δ::Hph/ura3Δ*); o número 2 indica a linhagem FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*); os números 3 e 4 são os dois transformantes da Etapa 04 (JAY270 *ura3Δ::CORE/ura3Δ*). A banda de 1900pb (JAO260/JAO14) corresponde à presença do cassete CORE no genoma.

Foi possível observar que o cassete CORE, embora ausente na linhagem “parental” FGY030 (1 – Figura 09), estava presente nos transformantes (3 e 4 – Figura 09), devido a amplificação da banda esperada de 1900pb.

A reação de PCR com os *primers* JAO260 e JAO14 seguiu as mesmas condições utilizadas na confirmação da Etapa 02.

Para corroborar o término da Etapa 04, a linhagem FGY040 (JAY270

ura3Δ::CORE/ura3Δ) passou por testes genotípicos e fenotípicos, a fim de verificar a presença e ausência das marcas genéticas nos seus esporos. Para isso, a linhagem diplóide foi submetida inicialmente à esporulação e dissecação de tétrades.

Os resultados dos dois testes envolvendo os quatro esporos da FGY040 foram confrontados. Dois esporos de uma tétrade, incapazes de crescer em quaisquer meios seletivos (YEPD+Higromicina; YEPD+G418 e YNB sem uracila), amplificaram uma banda de 144pb (JAO259/JAO260), referente à ausência de marcas genéticas no *locus URA3* (*ura3Δ*). Por outro lado, os outros dois esporos da mesma tétrade, que amplificaram uma banda de aproximadamente 3200pb relativa à presença do cassete CORE (JAO259/JAO260), foram capazes de crescer em meio contendo G418 e em meio sem uracila, sendo sensíveis à droga higromicina (resultados não mostrados).

Ao término desta etapa, a linhagem FGY040, recuperou a condição *URA+*, dada a presença do gene *κIURA3* ativo no seu genoma, estando, assim, pronta para passar pela quinta e última etapa do processo de domesticação.

Vale lembrar mais uma vez, que a condição *URA+* foi essencial para que a marca genética de seleção possa ser retirada com o fragmento homólogo de reparo. É exatamente por isso que a etapa 04 está presente no processo de domesticação da JAY270. Se esta etapa não fosse crucial não haveria a necessidade de retirar o cassete Hph pela introdução do cassete CORE, para depois transformar com o fragmento homólogo de reparo.

A última etapa do processo de domesticação da linhagem JAY270 consistiu na retirada das marcas genéticas *kanMX4* e *κIURA3* da linhagem FGY040 (*JAY270 ura3Δ::CORE/ura3Δ*) utilizando o fragmento homólogo de reparo.

Sabia-se, previamente, que esta etapa apresentaria uma dificuldade semelhante à deparada na etapa 03, com relação à procura pelo transformante de interesse. Isto, pois ao fazer a seleção em meio contendo 5-FOA também haveria o crescimento de leveduras não-transformadas, devido aos três eventos elucidados anteriormente, que as tornariam auxotróficas para uracila e, portanto, não susceptíveis ao composto tóxico: (I) mutação natural no gene *κIURA3*; (II) inativação da transcrição do *κIURA3* e (III) recombinação mitótica, segregando parte do braço esquerdo dos cromossomos V em homozigose (*ura3Δ/ura3Δ*).

Ao fazer a seleção negativa dos candidatos em meio contendo geneticina, os mutantes naturais (I) e os decorrentes da inativação da transcrição (II) seriam facilmente eliminados, devido à presença da marca genética *kanMX4* no seu genoma. Dessa forma, a seleção final se concentraria na distinção da levedura transformada FGY050 (*JAY270 ura3Δ/ura3Δ*) dentre os

recombinantes mitóticos (JAY270 *ura3Δ/ura3Δ*).

Entretanto, neste caso os *primers* JAO259/JAO260 (Figura 04) não seriam capazes de tal diferenciação, tendo em vista que tanto os transformantes de interesse quanto os recombinantes apresentariam o mesmo padrão genotípico para o *locus URA3* (144pb; *ura3Δ*).

A única diferença existente entre a levedura transformada FGY050 e os recombinantes mitóticos estaria localizada no braço esquerdo do cromossomo V, mais especificamente do *locus URA3* em direção ao telômero esquerdo (TEL05). Enquanto a linhagem JAY270 exibe um genoma altamente heterozigoto, os recombinantes passariam a apresentar homozigose nesta parte do braço esquerdo do cromossomo V (TEL05 ao *locus URA3*), decorrente do evento de recombinação mitótica, tal como constatado durante a etapa 03. Sabendo-se disso, algumas regiões heterozigotas localizadas no braço esquerdo do cromossomo V da JAY270 foram identificadas baseando-se no seqüenciamento do genoma do esporo JAY291 do trabalho de Argueso *et al.*, 2009, Genome Research, 19(12): 2258-70, incorporado aqui sua referência, e na montagem de alguns *contigs* do esporo JAY292 (ambos esporos de uma mesma tétrade da JAY270).

Primers foram desenhados conforme mostrado na Tabela 03, para amplificar estas regiões de heterozigosidade da JAY270, de modo que enzimas de restrição específicas pudessem reconhecer os polimorfismos *SNPs* (*Single Nucleotide Polymorphism*), gerando um padrão genotípico diferente para cada uma das cópias dos esporos JAY291 e JAY292. Esta técnica de detecção de polimorfismos usando enzimas de restrição é conhecida como RFLP (*Restriction Fragment Length Polymorphism*). Dessa forma, o padrão de digestão enzimática do transformante desejado seria diferente para cada uma das cópias que compõem a levedura diplóide, enquanto que um mesmo padrão genotípico de digestão seria observado para ambas às cópias do recombinante mitótico, devido à segregação em homozigose e, assim, à ausência do polimorfismo.

Tabela 03 - Sequência dos *primers* JAO311 e JAO312, desenhados a confirmação da Etapa 05 da domesticação da JAY270. Eles estão localizados na região 90Kb do braço esquerdo do cromossomo V da levedura, estando o gene *URA3* na posição 116-117Kb.

<i>Primer</i>	SEQ ID	Sequência (5' → 3')
JAO311	13	GGCAATGTCTCTTCCACCAT
JAO312	14	CAATTCGTTGTTTGCAATG

A linhagem FGY040 (JAY270 *ura3Δ::CORE/ura3Δ*) foi transformada com o fragmento homólogo de reparo e, após fazer a segunda seleção em meio com 5-FOA, os mutantes naturais e os decorrentes de inativação da transcrição foram eliminados através da seleção negativa em meio contendo geneticina (G418).

Feito isso, os candidatos sensíveis a G418 e incapazes de crescer em meio sem uracila foram submetidos inicialmente à reação de PCR, utilizando-se os *primers* JAO311/JAO312 para a amplificação do fragmento a ser digerido enzimaticamente, de 914pb (Figura 10). A localização dos *primers* JAO311/JAO312 foi baseada na sequência virtual do genoma da linhagem S288c (<http://www.yeastgenome.org/>) conforme apresentado na Figura 10 (a) e Figura 10 (b).

A figura 10(a) mostra a ilustração do braço esquerdo dos dois cromossomos V da levedura JAY270, estando representados pelos esporos JAY291 e JAY292. A estrela cinza corresponde aos pontos de restrição da enzima BslI indicando a existência de polimorfismos *SNP* entre os dois esporos de uma mesma tétrade da levedura diplóide e a Figura 10(b) após a digestão enzimática a JAY270 passa a apresentar um padrão genotípico (150pb/353pb/411pb/503pb) que indica a presença de heteroziguidade. Legenda: CEN5 – centrômero do cromossomo V; TEL05L – telômero esquerdo do cromossomo V.

Embora as figuras 10 (a) e 10 (b) estejam fora de escala, pode-se verificar que estes *primers* se anelam em um local específico do braço esquerdo do cromossomo V da levedura JAY270, entre o gene *URA3* (posição 116 - 117Kb do genoma da S288c) e o telômero esquerdo, TEL05L (posição 0Kb).

Na figura 10 (a) tem-se a localização dos *primers* JAO311 e JAO312 no genoma da JAY270, sendo possível identificar a existência de polimorfismo entre os esporos JAY291 e JAY292 devido ao número diferente de pontos de restrição para a enzima BslI (*).

Na figura 10 (b) tem-se o padrão genotípico esperado de ambos os esporos da JAY270 resultante da digestão da enzima BslI.

Para determinar o padrão genotípico dos candidatos e, assim, identificar os transformantes de interesse dentre os recombinantes mitóticos, inicialmente efetuou-se a amplificação do fragmento de interesse (914pb) com os *primers* JAO311 e JAO312. Estão representados 10 dos 26 candidatos testados conforme mostrado na Figura 11, onde M é o marcador molecular (1Kb plus invitrogen); os números 1 a 10 indicam os dez dos 26 candidatos testados.

A reação de PCR seguiu as seguintes condições: 1µl de DNA, 3µl de tampão (10X),

1,5µl de MgCl₂ (50mM), 0,42µl de cada um dos *primers* (15pmol), 6µl de dNTPs (1,25mM) e 0,6µl de taq DNA polimerase (50U/µl) (*Invitrogen*) para um volume final de 30µl. O programa utilizado foi: um ciclo inicial de desnaturação a 94°C por 2 minutos, seguido por 40 ciclos de 94°C por 30 segundos, 44 a 53°C por 30 segundos e 72°C por 1 minuto, além da extensão final de 72°C por 5 minutos. Os produtos de PCR foram visualizados em gel de agarose (1,5 a 2%) por um período de 1 a 3 horas, utilizando-se corrente entre 60 a 70V e, visualizados em fotodocumentador após marcação com brometo de etídeo.

Uma vez amplificado, o fragmento de 914pb foi submetido à digestão enzimática.

Conforme mencionado anteriormente na figura 10, existe dois pontos de restrição para a enzima BslI na linhagem JAY270. Um ponto de restrição era comum a ambas as cópias da JAY291 e da JAY292, estando localizado na posição 411 (411pb e 503pb). O outro ponto de restrição estava presente apenas na sequência da JAY291, na posição 561. Assim, enquanto o padrão genotípico da JAY292 era constituído por apenas duas bandas (411pb e 503pb), o da JAY291 foi composto por três bandas (150pb, 353pb e 411pb). Logo, o padrão genotípico do transformante FGY050 é formado pelas quatro bandas (150pb, 353pb, 411pb e 503pb), dada a existência do polimorfismo e, assim, dos dois pontos de restrição. Qualquer um dos dois padrões genotípicos (411pb/503pb ou 150pb/353pb/411pb) diferentes deste (150pb/353pb/411pb/503pb) indicaria a ausência do polimorfismo e, portanto, a ausência de heterozigosidade dos recombinantes mitóticos mostrado na Figura 10 (b).

Após a amplificação do fragmento de 914pb, o produto de PCR foi submetido à digestão enzimática com a enzima de restrição BslI.

A digestão foi feita seguindo o protocolo de fabricação (*Fermentas*): 20µL do produto de PCR + 4µL do tampão B (10X) da enzima + 1µL da enzima BslI + 15µL de água, para um volume final de 40µL. Esta mistura foi incubada a 55°C durante 16 horas.

Os subprodutos da digestão foram visualizados em gel de agarose (1,5 a 2%) por um período de 1 a 3 horas, utilizando-se corrente entre 60 a 70V e, visualizados em fotodocumentador após marcação com brometo de etídeo.

Dos 26 candidatos, seis eram os transformantes de interesse (FGY050). Pode-se observar que dentre os 10 candidatos representados conforme mostrado na Figura 12, onde M é o marcador molecular (1Kb plus invitrogen); os números 1 a 10 indicam os dez dos 26 candidatos.

O perfil genotípico composto pelas quatro bandas (150pb; 353pb; 411pb e 503pb) é o esperado para o transformante FGY050, indicando a presença do polimorfismo e a permanência

da heterozigiosidade da JAY270 dois eram os transformantes desejados (2 e 10), já que mostraram um padrão genotípico composto pelas quatro bandas (150pb/353pb/411pb/503pb) que indicavam a presença de polimorfismo. Por outro lado, o perfil genotípico de todos os recombinantes (1; 3 a 9) era constituído por três bandas (150pb/353pb/411pb). Isto indica que a marca *ura3Δ::CORE* da linhagem FGY040 estava na cópia da JAY292, já que parte do braço esquerdo do cromossomo V da JAY291, incluindo a segunda cópia do *locus URA3 (ura3Δ)*, segregou em homozigose durante a recombinação mitótica, carregando consigo os dois pontos de restrição para a enzima *e*, por isso, gerando as três bandas.

Para corroborar o término da etapa 05 e, portanto, do processo de domesticação da JAY270, um segundo teste foi realizado para que fosse possível observar o padrão genotípico dos esporos JAY291 e JAY292, verificar a existência de heterozigiosidade na linhagem JAY270, bem como constatar a permanência de heterozigiosidade nos transformantes FGY050. Os *primers* JAO311 e JAO312 foram utilizados para amplificar o fragmento de 914pd de diferentes linhagens como mostrado na Figura 13, onde M é o marcador molecular (1Kb plus invitrogen); o número 1 é JAY291; o número 2 é JAY292; o número 3 é JAY270; o número 4 é FGY040; os números 5 e 6 são os transformantes FGY050 e os números 7 e 8 são os recombinantes mitóticos.

A figura 13 serviu de base para a digestão enzimática subsequente com a enzima *BslI* conforme mostrado na Figura 14, onde M é marcador molecular (1Kb plus invitrogen); os números: 1 é JAY291; 2 é JAY292; 3 é JAY270; 4 é FGY040; 5 e 6 são transformantes FGY050; 7 e 8 são recombinantes mitóticos.

O padrão genotípico da JAY270 (150pb; 353pb; 411pb e 503pb) foi verificado para os dois transformantes FGY050 (5 e 6), indicando a permanência da heterozigiosidade ao final do processo de domesticação.

Ficou evidente que as leveduras transformantes FGY050 (5 e 6 - Figura 14) mantiveram a heterozigiosidade da linhagem JAY270 selvagem (3 – Figura 14) e da levedura FGY040 da etapa 04 (4 – Figura 14), já que todas mostraram o mesmo padrão genotípico com as quatro bandas esperadas (150pb, 353pb, 411pb e 503pb).

Pode-se notar que o padrão de bandas delas é uma complementação dos padrões genotípicos dos esporos JAY291 e JAY292 (1 e 2, respectivamente).

A JAY291 apresentou um padrão genotípico com três bandas (150pb/353pb/411pb), indicando que existem realmente dois pontos de restrição dentro do fragmento de 914pb (posição 411 e 561).

Por outro lado, a JAY292 mostrou apenas duas bandas (411pb/503pb). Isto deixa claro que o fragmento de 914pb da JAY292 possui apenas um sítio de restrição (posição 411) para a enzima, sendo por esse motivo que o fragmento maior (503pb) não foi cortado. Por fim, foi possível visualizar também que os recombinantes mitóticos (7 e 8 – Figura 14) obedeciam o padrão genotípico da JAY291, comprovando, assim, a segregação em homozigose da segunda cópia do *locus URA3*, representada pelo *ura3Δ* “limpo”.

Estes resultados confirmam o término da quinta e última etapa do processo de domesticação da linhagem JAY270.

A manipulação genética desta linhagem diplóide foi alcançada com sucesso, sendo que a linhagem final FGY050 passou a apresentar a auxotrofia para uracila após a deleção completa das duas cópias do gene *URA3* do seu genoma.

Sequenciamento do *locus URA3* da linhagem FGY050 para confirmação da deleção completa das duas cópias do gene *URA3* da linhagem JAY270 efetuada na presente invenção.

Foram realizadas quatro reações independentes (réplicas) de seqüenciamento. Cada reação de seqüenciamento contou com 2µl de Big Dye, 0.5µl de *primer* 5pmol (JAO259 ou JAO260), 0.2 a 0.8µl de DNA (Produto de PCR) e água Milli-Q suficiente para completar 10µl de reação. O programa consistiu de 2 minutos iniciais a 94°C, 35 ciclos compreendendo 20 segundos a 94°C, 15 segundos a 57°C (temperatura de anelamento do *primer*) e 1 minuto a 60°C. Após a reação, o produto foi estocado sob refrigeração à 4°C. Os dNTPs não incorporados foram eliminados por precipitação com álcool isoamílico 75% e etanol 70% gelado. Após a secagem completa do DNA, acrescentou-se 10µl de solução de formamida. Depois de ressuspendido, a solução com DNA foi desnaturada a 95°C durante 5 minutos e imediatamente imersa em gelo até o momento do seqüenciamento, que foi realizado em seqüenciador 3500 *Genetic Analyzer*, conforme protocolo do fabricante (*Applied Biosystems*).

Antes da reação de seqüenciamento descrita acima foi realizada uma reação de PCR, cujo produto foi usado como DNA para o próprio seqüenciamento. Esta reação de PCR foi feita segundo as seguintes condições: 2µl de DNA genômico da linhagem FGY050, 2µl de tampão (10X), 0,6µl de MgCl₂ (50mM), 0,28µl de cada um dos *primers* (JAO259 e JAO260) (15pmol), 4µl de dNTPs (1,25mM) e 0,4µl de taq DNA polimerase (50U/µl) (*Invitrogen*) para um volume final de 20µl. O programa utilizado foi: um ciclo inicial de desnaturação a 94°C por 2 minutos, seguido por 35 ciclos de 94°C por 30 segundos, 57°C por 30 segundos e 72°C por 1 minuto,

além da extensão final de 72°C por 5 minutos.

Análise das Sequências

As sequências foram analisadas pelos programas Phred/Phrap/Consed (Ewing *et al.*, 1998; Gordon, Desmarais & Green, 2001). Somente sequências com ao menos 200 bases com qualidade Phred igual ou maior a 20 (probabilidade de um erro para cada 1000 bases) foram utilizadas. As sequências com boa qualidade foram alinhadas pelo programa ClustalX e agrupadas, formando um único conjunto de interesse. Esta sequência foi submetida ao banco de dados GenBank do NCBI (<http://www.ncbi.nlm.nih.gov/>) e do SGD (<http://www.yeastgenome.org/>) para a busca de similaridade frente a sequências já depositadas, através do programa Blastn.

Resultado

I - Sequência alinhada no ClustalX:

```
ATGTGGTGTTGAAGAAACATGAAATTGCCAGTATTCTTAACCCAACCTGCACAGAACAAAAACCTGCATGCATGTATA  
CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCGGGAATCTCGGTCGT
```

Tamanho do fragmento sequenciado: 141pb

II - Similaridade com o banco de dados do SGD:

Na figura 15 está representado parte do Cromossomo V da levedura *Saccharomyces cerevisiae*, segundo o banco de dados SGD. Pode-se observar que o gene *URA3* (YEL021W) está localizado dentro das posições 116K e 117K, conforme mostrado na figura 15.

A região de DNA sequenciada a partir da levedura modificada FGY050 teve 100% de similaridade com a sequência virtual da levedura *S. cerevisiae* do SGD.

Uma parte do DNA sequenciado (68 bases iniciais; sublinhadas) mostrou homologia com a sequência destacada pelo retângulo tracejado (extremo 116K - *upstream* do locus *URA3*), enquanto que a outra parte do DNA sequenciado (73 bases finais; não sublinhadas) teve homologia com a sequência destacada pelo retângulo de linhas contínuas (extremo 117K – *downstream* do locus *URA3*) conforme observado na Figura 16.

```
ATGTGGTGTTGAAGAAACATGAAATTGCCAGTATTCTTAACCCAACCTGCACAGAACAAAAACCTGCATGCATGTATA  
CTAAACTCACAAATTAGAGCTTCAATTTAATTATATCAGTTATTACCCGGGAATCTCGGTCGT
```

Pode-se observar pela Figura 16 que o fragmento de DNA sequenciado a partir da linhagem FGY050 (141pb) mostrou similaridade somente com as regiões flangeadoras do locus *URA3* (destacadas pelos retângulos de linhas tracejadas e contínuas). Por outro lado, o mesmo DNA sequenciado não mostrou similaridade com a sequência do gene *URA3*.

Esse resultado comprova que a sequência do gene *URA3* não está presente na

linhagem FGY050, corroborando a deleção completa das duas cópias do gene, que foi recém efetuada no presente invento.

Se os *primers* JAO259 e JAO260 fossem usados no sequenciamento da sua parental JAY270, haveria a amplificação de toda a sequência destacada entre os retângulos de linhas tracejadas e contínuas, incluindo o gene *URA3*. Com isso, a similaridade entre o DNA sequenciado a partir da linhagem JAY270 e o banco de dados da *Saccharomyces cerevisiae* resultaria em apenas uma única sequência conforme mostrado na Figura 17 e não em duas partes diferentes (sendo uma na região *upstream* e outra na região *downstream* do locus *URA3*) como foi observado para o sequenciamento da FGY050.

O resultado, aqui obtido, mostra que a linhagem modificada neste invento (FGY050) possui um genoma diferente da sua linhagem parental JAY270, que é uma levedura selvagem (prototrófica, ou seja, sem nenhuma marca de auxotrofia). Enquanto a JAY270 possui as duas cópias do gene *URA3* ativas no seu genoma prototrófico, a linhagem FGY050 não possui quaisquer cópias deste gene no seu genoma e, por isso, ela é auxotrófica para uracila, já que teve esta via metabólica interrompida.

Além de conseguir a auxotrofia desejada para a JAY270, a manipulação genética realizada através do sistema *Delitto Perfetto* permitiu ainda preservar o genoma altamente heterozigoto da linhagem JAY270, e assim, manter as suas características de rusticidade e robustez de interesse fermentativo industrial.

Assim, a levedura manipulada da presente invenção FGY050 é geneticamente igual à sua parental selvagem, exceto pela deleção das duas cópias do gene *URA3* que lhe conferiu fenotipicamente a incapacidade de crescer em meio sem uracila.

Assim sendo, ao preservar as características da linhagem JAY270 e, portanto, da sua cepa mãe PE-2, serão maiores as chances de sucesso, de linhagens a serem melhoradas geneticamente (a partir da FGY050), em futuras aplicações industriais. O processo de manipulação genética foi efetuado sem deixar quaisquer rastros de transgenia no genoma da linhagem modificada, o que é importante tendo em vista a reintrodução e, o futuro emprego, da linhagem manipulada nos processos industriais de fermentação.

Portanto, a levedura manipulada é geneticamente igual à sua parental selvagem JAY270 exceto pela auxotrofia gerada que resultou na sua incapacidade de crescer em meio sem uracila.

As considerações acima são vantagens atribuídas à presente invenção, que lhe garante maiores chances de sucesso em futuras aplicações industriais.

A presente invenção permitirá realizar inúmeros estudos de melhoramento genético usando um gene da própria natureza genética da levedura, o *URA3*, como marca de seleção dos transformantes melhorados. Com isso, genes bacterianos de resistência a drogas não precisarão ser utilizados como marcadores seletivos, não havendo problemas relacionados à produção ou manipulação de microrganismos transgênicos. De acordo com a literatura, o número de transformantes falso-positivos gerados pelo uso de genes marcadores auxotróficos é menor do que ao se usar genes bacterianos de resistência a drogas.

Além das etapas de deleção das duas cópias do *URA3* foram necessárias outras etapas até que as marcas *hphMX4* e *kanMX4* fossem retiradas do genoma da levedura JAY270. Esse cuidado foi super importante para que a levedura final modificada não apresentasse marcas de transgenia e, portanto, não fosse considerada como um microrganismo transgênico, ainda mais sabendo que em um futuro próximo ela será utilizada em estudos de melhoramento genético para ser empregada em aplicações industriais, com possíveis fins comerciais. Assim, nesse futuro próximo, não haverá a necessidade de uma usina ou empresa gastar com controles de disseminação de OGM transgênicos no meio ambiente, já que estará usando um organismo não transgênico.

Assim, embora tenham sido mostradas apenas algumas modalidades da presente invenção, será entendido que várias omissões, substituições e alterações podem ser feitas por um técnico versado no assunto, sem se afastar do espírito e escopo da presente invenção.

Também é preciso entender que os desenhos não estão necessariamente em escala, mas que eles são apenas de natureza conceitual. A intenção é, portanto, ser limitada, tal como indicado pelo escopo das reivindicações anexas.

REIVINDICAÇÕES

1. Processo para produção de uma levedura mutante nula do gene *URA3* da cepa industrial Pedra-2 de *Saccharomyces cerevisiae* **caracterizado pelo** fato de compreender as etapas de:
 01. Deleção completa de uma das duas cópias do gene *URA3* da levedura utilizando um cassete de integração;
 02. Deleção da segunda cópia do gene *URA3* da levedura FGY010 com o auxílio de um cassete de integração;
 03. Retirada do cassete CORE da 2ª cópia do *locus URA3* do genoma da levedura FGY020, utilizando o fragmento homólogo de reparo;
 04. Retirada do cassete Hph, contendo a marca genética *hphMX4*, do genoma da levedura FGY030, pela integração do cassete CORE;
 05. Retirada das marcas genéticas *kanMX4* e *κURA3* da linhagem FGY040, utilizando o fragmento homólogo de reparo.
2. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que o referido cassete de integração da etapa (01) é denominado “cassete Hph”, o qual possui a marca genética *hphMX4* de resistência à droga higromicina.
3. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a linhagem transformante na etapa (01) foi denominada como FGY010 e seu genótipo é igual ao da linhagem JAY270 exceto pela deleção de uma cópia do gene *URA3* (JAY270 *ura3Δ::Hph/URA3*).
4. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que o referido cassete de integração da etapa (02) é denominado “cassete CORE”, o qual possui o gene *kanMX4* de resistência à droga geneticina (G418) e, também, o gene *κURA3* da levedura *Kluyveromyces lactis*.
5. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a linhagem transformante na etapa (02) foi denominada como FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*).
6. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a linhagem transformante na etapa (02) apresenta marcas de resistência a dois antibióticos (higromicina e geneticina),
7. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a linhagem transformante na etapa (02) apresenta o gene

κIURA3 ativo no seu genoma.

8. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que o referido fragmento homólogo de reparo da etapa (03) é composto apenas pelas sequências das regiões upstream e downstream do gene *URA3*.

9. Processo para produção de uma levedura mutante, de acordo com a reivindicação 8, **caracterizado pelo** fato de que o referido fragmento homólogo de reparo da etapa (03) não apresenta nenhuma marca de seleção no seu interior.

10. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a linhagem transformante na etapa (03) foi denominada de FGY030 (JAY270 *ura3Δ::Hph/ura3Δ*).

11. Processo para produção de uma levedura mutante, de acordo com a reivindicação 10, **caracterizado pelo** fato de que a linhagem transformante na etapa (03) já apresenta auxotrofia para uracila, mas com uma marca de transgenia (*hphMX4*).

12. Processo para produção de uma levedura mutante, de acordo com a reivindicação 10, **caracterizado pelo** fato de que a seleção da linhagem transformante na etapa (03) é verificada pela perda do gene *κIURA3* em meio contendo 5-FOA.

13. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a etapa (04) é feita através da integração do cassete CORE com a linhagem FGY030.

14. Processo para produção de uma levedura mutante, de acordo com a reivindicação 13, **caracterizado pelo** fato de que na etapa (04), ao transformar a linhagem FGY030 com o cassete CORE, a levedura passa a ser capaz de crescer em meio sem uracila, dada a presença do *κIURA3*.

15. Processo para produção de uma levedura mutante, de acordo com as reivindicações 13 e 14, **caracterizado pelo** fato de que a linhagem transformante na etapa (04) é denominada de FGY040 = JAY270 *ura3Δ::CORE/ura3Δ*.

16. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que na etapa (05), a levedura mutante formada é isenta das marcas genéticas *kanMX4* e *κIURA3*.

17. Processo para produção de uma levedura mutante, de acordo com a reivindicação 1, **caracterizado pelo** fato de que a linhagem transformante formada na etapa (05) é denominada de FGY050 = JAY270 *ura3Δ/ura3Δ*.

18. Levedura mutante desenvolvida pelo processo conforme definida nas reivindicações 1

a 17 **caracterizado pelo** fato de ser a levedura FGY050 (JAY270 *ura3Δ/ura3Δ*).

19. Levedura mutante, de acordo com a reivindicação 18, **caracterizado pelo** fato de ser uma levedura diplóide.

20. Levedura mutante, de acordo com a reivindicação 18, **caracterizado pelo** fato de ser auxotrófica para uracila.

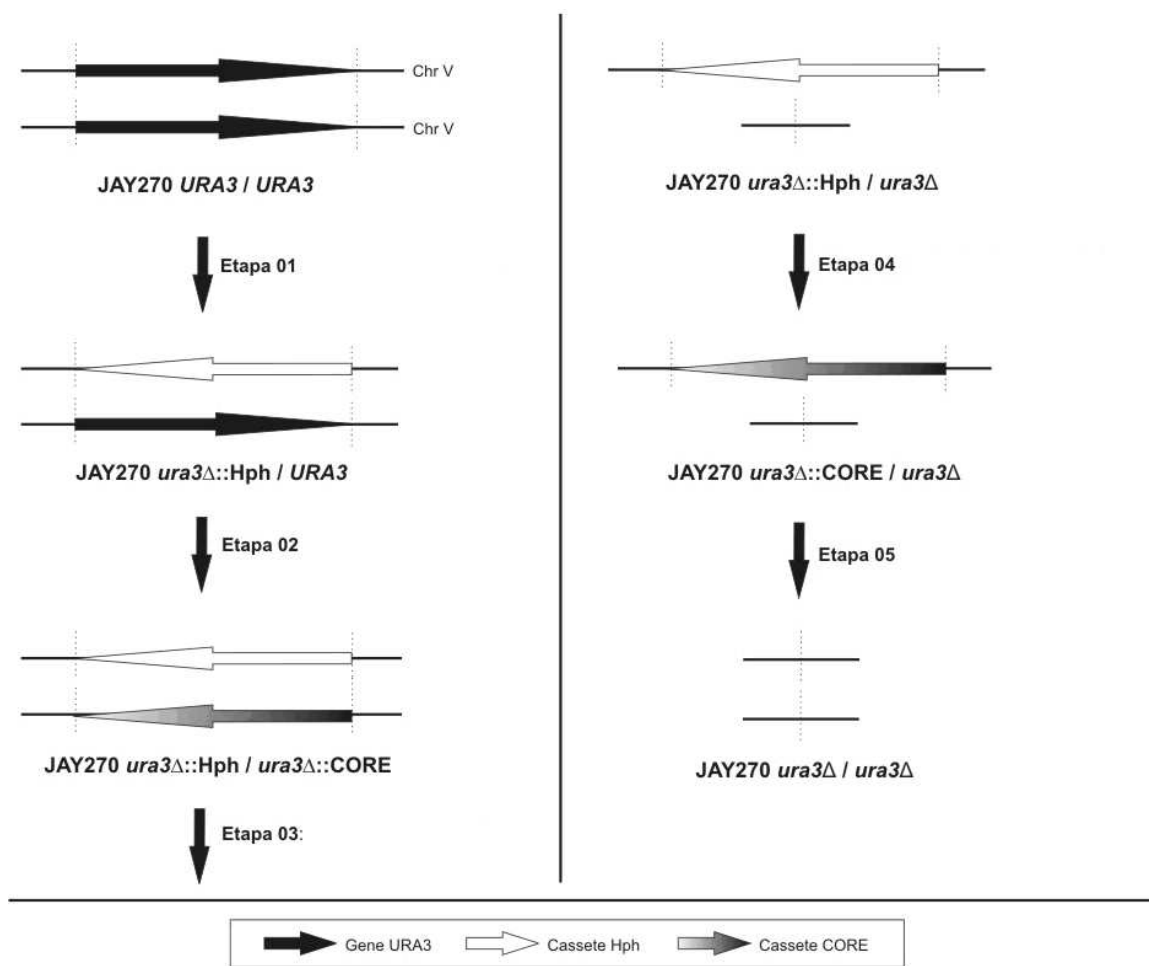
21. Levedura mutante, de acordo com a reivindicação 18, **caracterizado pelo** fato de que a região de DNA seqüenciada a partir da levedura modificada FGY050 possui 100% de similaridade com a sequência virtual da levedura *S. cerevisiae* do SGD.

22. Levedura mutante, de acordo com a reivindicação 18, **caracterizado pelo** fato de que possui alta resistência ao estresse.

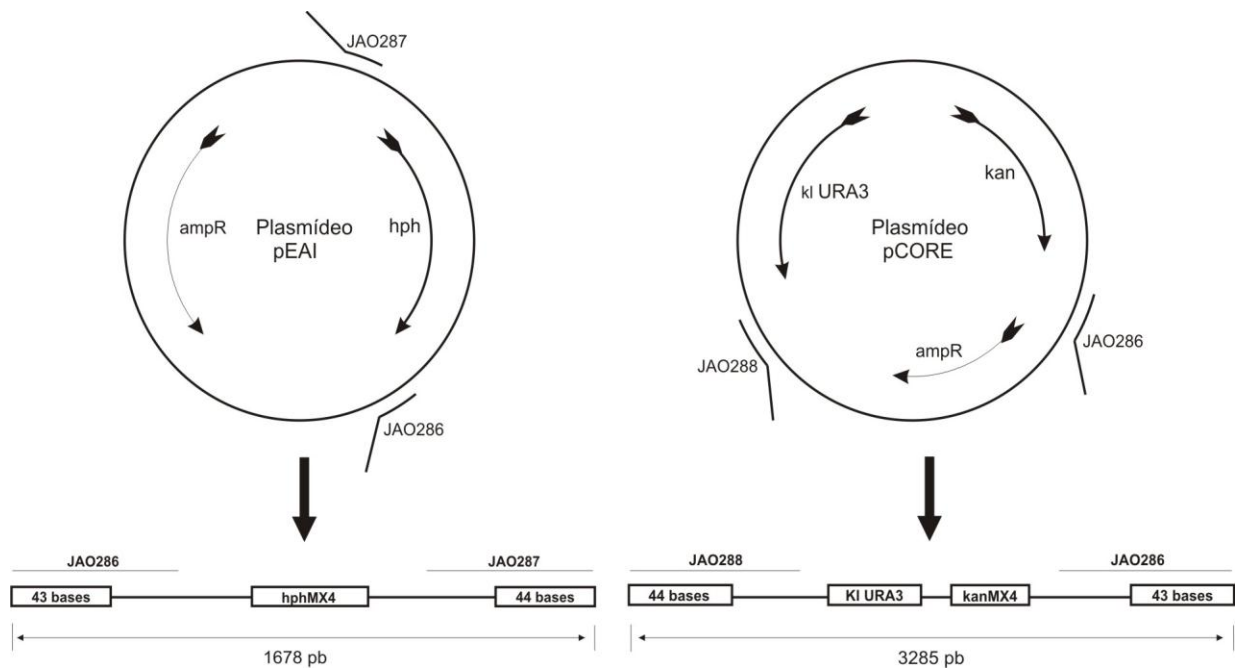
23. Levedura mutante, de acordo com a reivindicação 18, **caracterizado pelo** fato de que possui boa produtividade fermentativa.

24. Uso da levedura mutante conforme definida nas reivindicações 18 a 22 **caracterizado pelo** fato de ser para metabolizar fontes diversas para produzir diferentes metabólitos industriais selecionado do grupo compreendido por álcool, glicerol, eteno, propeno, biodiesel, hidrocarbonetos, proteínas específicas.

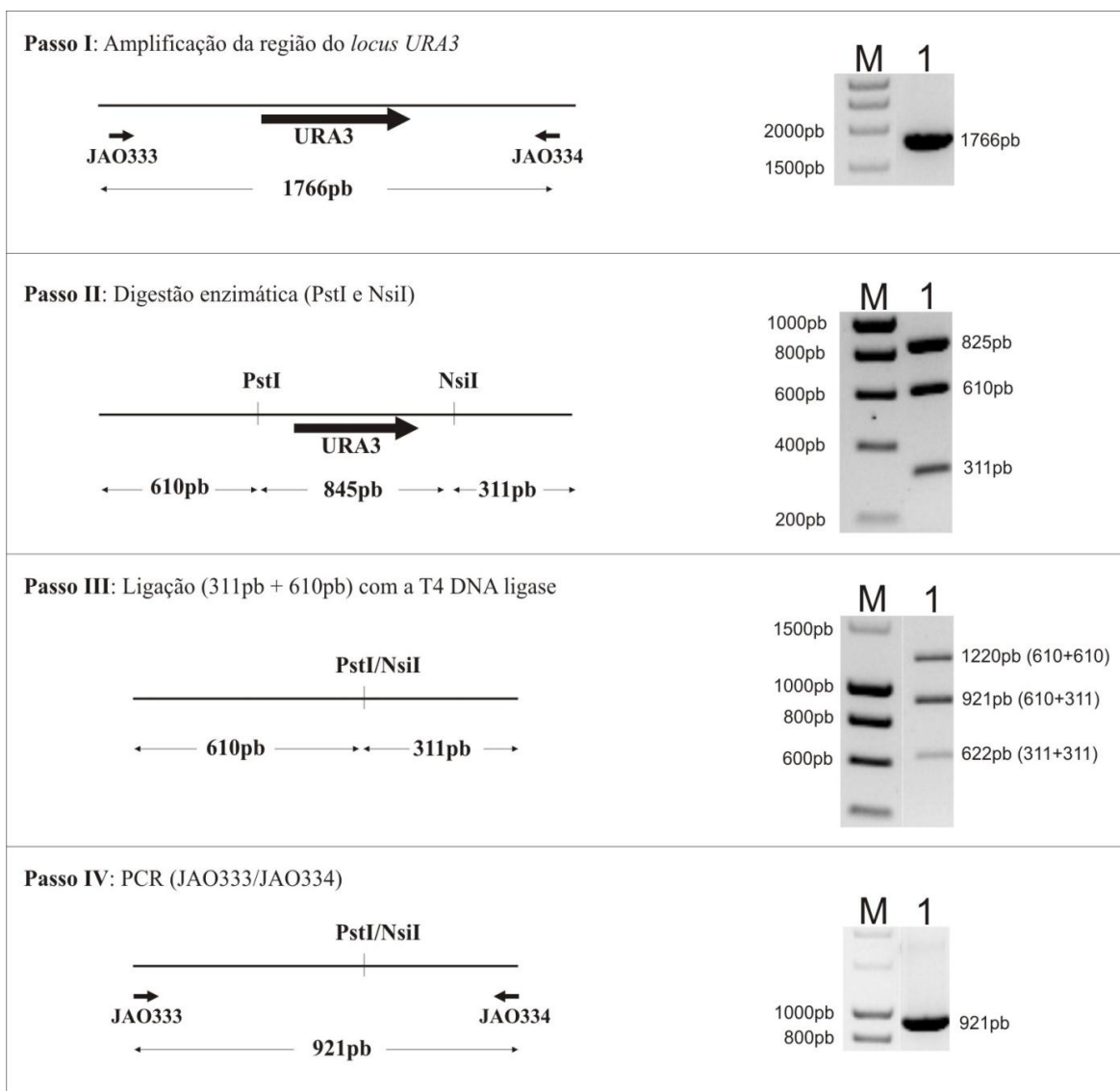
25. Uso da levedura mutante conforme definida nas reivindicações 18 a 22 **caracterizado pelo** fato de ser preferencialmente para produção de álcool a partir de cana-de-açúcar.



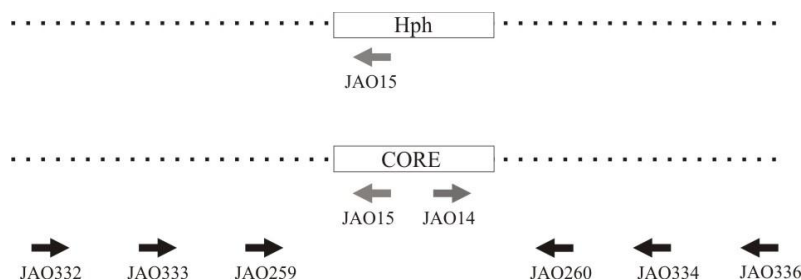
A Figura 1 apresenta o processo de domesticação da linhagem JAY270. Cinco etapas de manipulação genética foram necessárias para a deleção completa das duas cópias do gene *URA3* dos cromossomos V (Chr V) da levedura diplóide, sem deixar quaisquer marcas genéticas de seleção, utilizadas durante o processo, no genoma da levedura modificada FGY050 (JAY270 *ura3Δ/ura3Δ*), sendo que na etapa 01 ocorre a deleção da primeira cópia do gene *URA3* com o cassete *Hph*; na etapa 02 ocorre a deleção da segunda cópia do gene *URA3* com o casse *CORE*; na etapa 03 ocorre a retirada do cassete *CORE* com o fragmento homólogo de reparo; na etapa 04 ocorre a retirada do casse *Hph* com o cassete *CORE*; e na etapa 05 o cassete *CORE* é retirado com o fragmento homólogo de reparo.



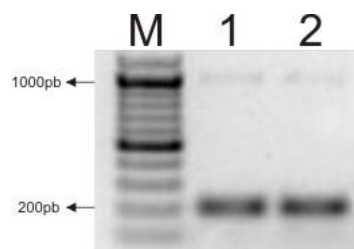
A Figura 2 apresenta a construção dos cassetes de integração (Hph e CORE) utilizados durante o processo de domesticação da linhagem JAY270. Diferentes combinações de *primers* (JAO286/JAO287 e JAO286/JAO288) permitiram a amplificação dos cassetes Hph e CORE a partir dos plasmídeos pEAI (similar ao pAG32) e pCORE, respectivamente. As extremidades dos cassetes são compostas por aproximadamente 45 bases contendo homologia com as regiões flanqueadoras do gene *URA3*.



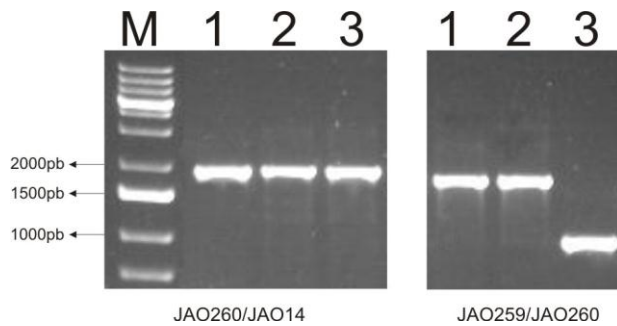
A Figura 3 apresenta a estratégia (lado esquerdo) referente à construção do fragmento homólogo de reparo, bem como os produtos de PCR, de digestão e de ligação (lado direito) resultantes. JAO333 e JAO334 são *primers* enquanto que PstI e NsiI são enzimas de restrição. M – marcador molecular (Bioline hyperladder I).



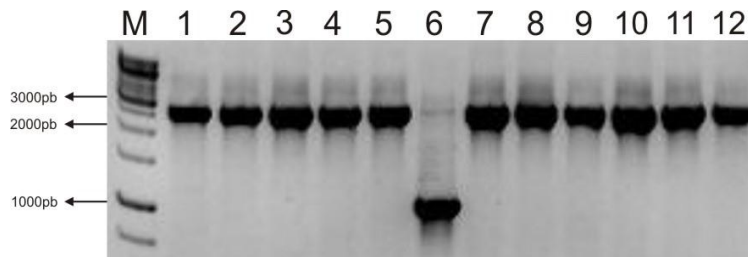
A Figura 4 apresenta a localização dos *primers* no genoma da levedura *Saccharomyces cerevisiae* utilizados para confirmar as transformações durante as etapas de domesticação da JAY270. Parte dos dois cromossomos V está representada na figura (em pontilhado) estando os cassetes Hph e CORE integrados no *locus URA3*. Os *primers* JAO259, JAO260, JAO333, JAO334, JAO332 e JAO336 estão localizados nas regiões *upstream* e *downstream* do gene *URA3* enquanto que os *primers* JAO14 e JAO15 são internos aos cassetes.



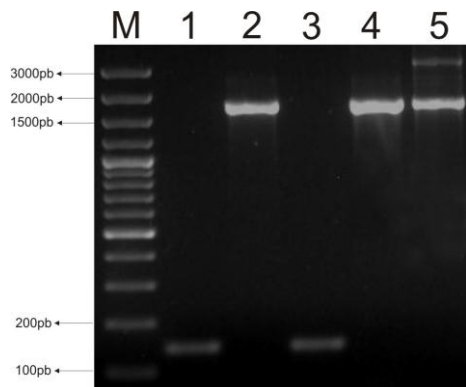
A Figura 5 apresenta os produtos de PCR referentes à confirmação do término da Etapa 01 do processo de domesticação da JAY270. M – marcador molecular (100pb); 1 e 2 – são os dois transformantes de interesse (FGY010). A banda de 1000pb corresponde à segunda cópia do gene *URA3* que se manteve intacta na linhagem transformante (JAY270 *ura3Δ::Hph/URA3*) enquanto que a banda de 200pb confirma a deleção da primeira cópia do gene *URA3*, decorrente da integração do cassete Hph.



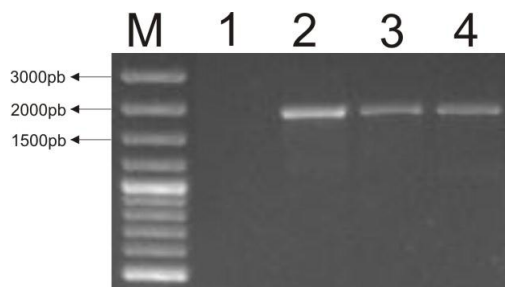
A Figura 6 apresenta os produtos de PCR referentes à confirmação do término da Etapa 02 do processo de domesticação da JAY270. M – marcador molecular (ladder 1Kb *Fermentas*); 1, 2 e 3 – transformantes candidatos. A banda de 1900pb (JAO260/JAO14) indica a integração do cassete CORE no genoma da levedura FGY010. Os *primers* JAO259/JAO260 mostram a permanência da marca *hphMX4* (1800pb) em apenas dois candidatos (1 e 2), enquanto que o outro candidato (3) manteve indesejadamente a segunda cópia do gene *URA3* intacta (1000pb).



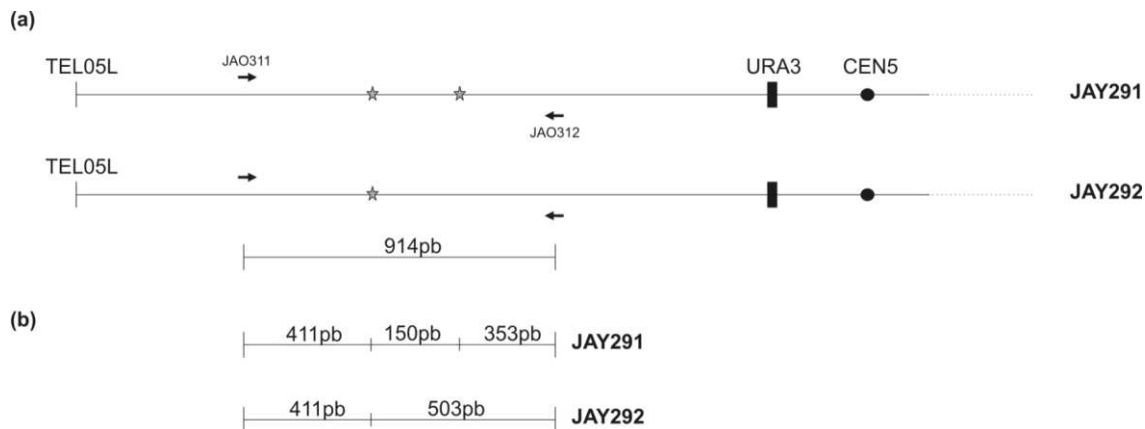
A Figura 7 apresenta os produtos de PCR referentes à confirmação do término da Etapa 03 do processo de domesticação da JAY270. M – marcador molecular (ladder 1Kb *Fermentas*); 1 a 11 – transformantes candidatos; 12 – controle positivo para a amplificação da marca *ura3Δ::Hph*. Somente o candidato 6 apresentou a banda esperada para o *ura3Δ* “limpo” (1100pb). A banda de 2700pb corresponde à marca *ura3Δ::Hph*. O candidato 6 é o próprio transformante FGY030 enquanto os outros são recombinantes mitóticos.



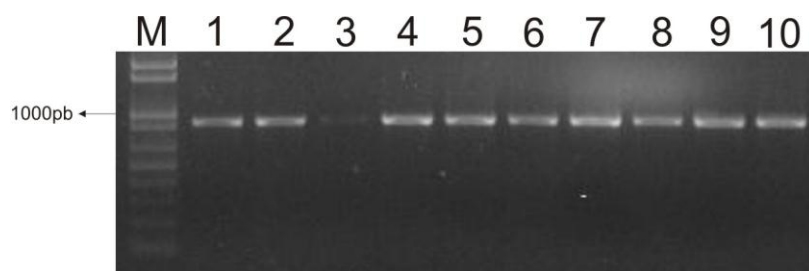
A Figura 8 apresenta os produtos de PCR dos quatro esporos de uma tétrade da linhagem FGY030, corroborando o término da Etapa 03 do processo de domesticação da JAY270. M – marcador molecular (ladder 100pb); 1 a 4 – os quatro esporos de uma tétrade da FGY030; 5 – linhagem “parental” FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*).



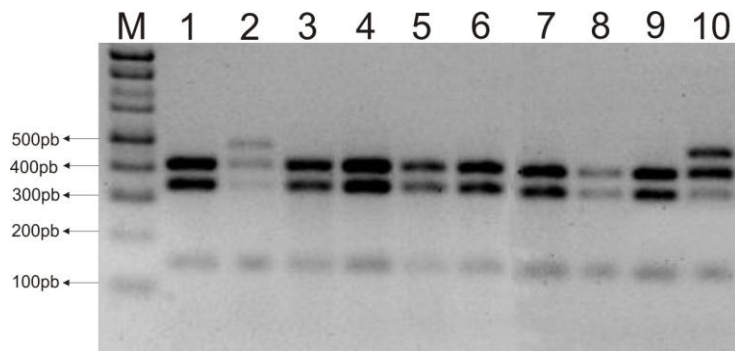
A Figura 9 apresenta os produtos de PCR referentes à confirmação do término da Etapa 04 do processo de domesticação da JAY270. M – marcador molecular (ladder 100pb); 1 – linhagem “parental” FGY030 (JAY270 *ura3Δ::Hph/ura3Δ*); 2 – linhagem FGY020 (JAY270 *ura3Δ::Hph/ura3Δ::CORE*); 3 e 4 – dois transformantes da Etapa 04 (JAY270 *ura3Δ::CORE/ura3Δ*). A banda de 1900pb (JAO260/JAO14) corresponde à presença do cassete CORE no genoma.



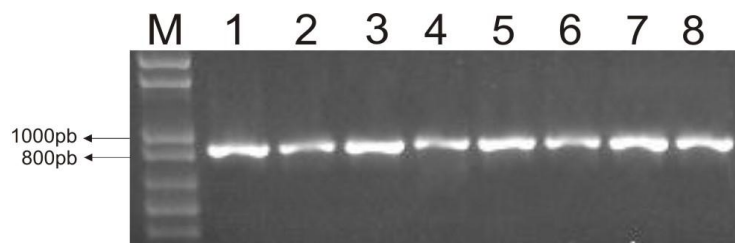
A Figura 10 apresenta o esquema da estratégia para a confirmação da Etapa 05 do processo de domesticação da JAY270. (a) Ilustração do braço esquerdo dos dois cromossomos V da levedura JAY270, estando representados pelos esporos JAY291 e JAY292. A estrela cinza corresponde aos pontos de restrição da enzima BslI indicando a existência de polimorfismos *SNP* entre os dois esporos de uma mesma tétrade da levedura diplóide. (b) Após a digestão enzimática a JAY270 passa a apresentar um padrão genotípico (150pb/353pb/411pb/503pb) que indica a presença de heteroziguidade. Legenda: CEN5 – centrômero do cromossomo V; TEL05L – telômero esquerdo do cromossomo V. A figura está fora de escala.



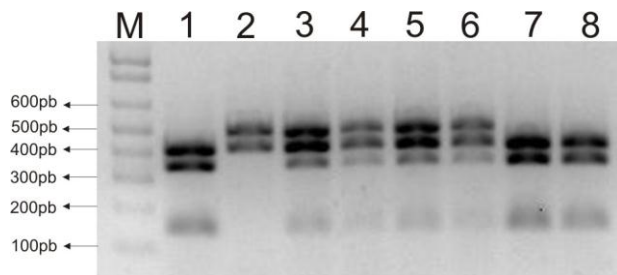
A Figura 11 apresenta os produtos de PCR (JAO311/JAO312) referentes à confirmação da Etapa 05 do processo de domesticação da JAY270. M – marcador molecular (1Kb plus invitrogen); 1 a 10 – dez dos 26 candidatos testados.



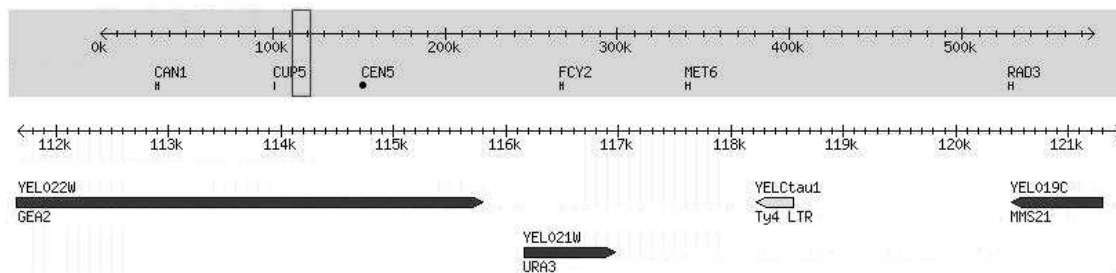
A Figura 12 apresenta os produtos de digestão da PCR referentes à confirmação da Etapa 05 do processo de domesticação da JAY270. Gel de agarose da digestão do produto de PCR (JAO311/JAO312) de 914pb (Figura 11) de 10 dos 26 candidatos testados. M – marcador molecular (1Kb plus invitrogen); 1 a 10 – dez dos 26 candidatos. O perfil genotípico composto pelas quatro bandas (150pb; 353pb; 411pb e 503pb) é o esperado para o transformante FGY050, indicando a presença do polimorfismo e a permanência da heterozigosidade da JAY270.



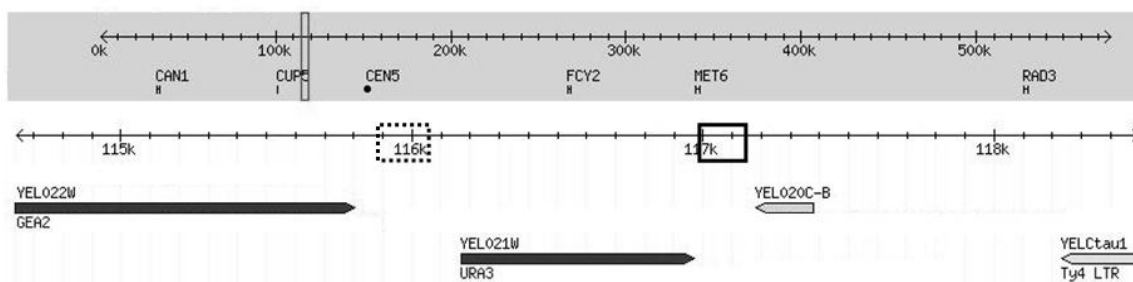
A Figura 13 apresenta os produtos de PCR para corroborar a permanência de heterozigosidade na linhagem FGY050 e, assim, o término da Etapa 05 do processo de domesticação da JAY270. Gel de agarose do produto de PCR (914pb), utilizando os *primers* JAO311 e JAO312. M – marcador molecular (1Kb plus invitrogen); 1 – JAY291; 2 – JAY292; 3 – JAY270; 4 – FGY040; 5 e 6 – transformantes FGY050; 7 e 8 – recombinantes mitóticos.



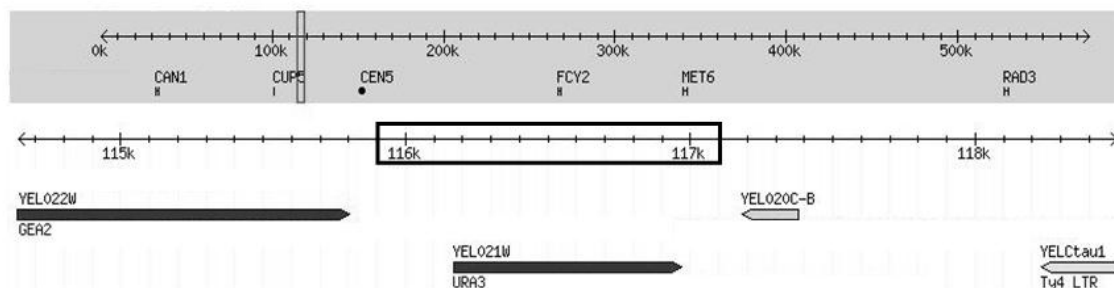
A Figura 14 apresenta o gel de agarose da digestão do fragmento de 914pb (JAO311/JAO312) obtido previamente (Figura 13). M – marcador molecular (1Kb plus Invitrogen); 1 – JAY291; 2 – JAY292; 3 – JAY270; 4 – FGY040; 5 e 6 – transformantes FGY050; 7 e 8 – recombinantes mitóticos. O padrão genotípico da JAY270 (150pb; 353pb; 411pb e 503pb) foi verificado para os dois transformantes FGY050 (5 e 6), indicando a permanência da heterozigosidade ao final do processo de domesticação da JAY270.



A Figura 15 é uma representação virtual de parte do Cromossomo V da levedura *Saccharomyces cerevisiae*, segundo o banco de dados SGD. O gene URA3 (YEL021W) está dentro das posições 116K e 117K.



A Figura 16 é uma representação virtual de parte do Cromossomo V da levedura *Saccharomyces cerevisiae* (SGD). A região de similaridade entre o DNA sequenciado da linhagem FGY050 e a sequência virtual da *S. cerevisiae* está destacada nos retângulos de linha tracejada e de linha contínua. O desenho está fora de escala.



A Figura 17 é uma representação virtual de parte do Cromossomo V da levedura *Saccharomyces cerevisiae* (SGD). No sequenciamento da região do locus *URA3* da linhagem selvagem JAY270 seria esperado apenas 01 único fragmento grande de similaridade, representado pelo retângulo de linha tracejada.

CAPÍTULO 4

***Saccharomyces cerevisiae* transcriptional reprogramming due to bacterial co-aggregation during an industrial Brazilian bioethanol production**

Osmar V. Carvalho-Netto ^{1,2}; Marcelo F. Carazzolle ¹; Luciana S. Mofatto¹; Paulo J.P.L. Teixeira ¹; Melline F. Noronha¹; Luige A.L. Calderón ¹; Piotr A. Mieczkowski ³; Juan Lucas Argueso ²; and Gonçalo A. G. Pereira ^{1,*}

¹ Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas-SP, Brazil.

² Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins-CO, USA.

³Department of Genetics, University of North Carolina, Chapel Hill-NC, USA.

* Corresponding author.

Corresponding author:

Gonçalo A. G. Pereira, PhD

Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas-SP, CEP: 13083-970, Brazil.

Phone: 55-19-3521-6237; Fax: 55-19-3521-6235; email: goncalo@unicamp.br

Background

The bioethanol production system used in Brazil is based on the alcoholic fermentation of sucrose derived from sugarcane feedstock by highly adapted strains of the yeast *Saccharomyces cerevisiae*. Due to no sterile condition at industrial scale, the process carries a variety of bacterial contaminants that are regularly related to yeast-bacteria co-aggregation phenotype, decreasing bioethanol yield.

Methodology and Principal Findings

In this study we investigate the molecular physiology of the main *S. cerevisiae* commercial strain (PE-2) used on Brazilian bioethanol process under two distinct conditions: typical (TF) and flocculated (co-aggregated - FL) fermentation. Transcriptional machinery of PE-2 was assessed by high throughput sequencing-based methods (RNA-seq) during industrial fed-batch fermentations. Data from comparative analysis revealed distinct transcriptional profiles among conditions, characterized mainly by a deep gene repression on FL process. We have showed that *Lactobacillus fermentum* were responsible for the co-aggregation phenotype and probably for the high levels of organic acids identified on samples. Moreover, we also detected by differential allelic expression an event of loss of heterozygosity at chromosome 13.

Conclusions

We have reported for the first time an overview of the *S. cerevisiae* transcriptional machinery during industrial bioethanol production in Brazil. The deep transcriptional reprogramming observed at flocculated condition indicated that both organic acids produced by *L. fermentum* and its physical interaction with yeast cell wall were highly detrimental to the process. Also, we provided a resource data regarding the mechanisms used by PE-2 during industrial fermentation with potential to be applied on further genetic improvements programs.

Keywords

Industrial fermentation, bioethanol, co-aggregation, transcriptome, RNA-seq.

INTRODUCTION

Brazilian bioethanol is produced by sugarcane juice and molasses fermentation carried out mainly by *Saccharomyces cerevisiae*. Yeasts inoculated at the beginning of the season are recycled after each fed-batch fermentation cycle (8-15 hours) during ~210 days. Since sugarcane juice and molasses are not sterile, they carry a variety of contaminants to the fermentation, resulting in a dynamic competition between the strain inoculated, wild yeast strains and bacteria [1,2].

PE-2 and CAT-1 are the most productive and widely adopted *S. cerevisiae* strains used by distilleries in Brazil [1]. We have recently described the complex genome structure of JAY291/PE-2 strain providing insights that explain its elevated fitness [3]. PE-2 is a wild type diploid strain, heterothallic and its genome is characterized by a high degree of heterozygosity, both at the nucleotide sequence and karyotype levels. This intrinsic genetic diversity is likely a key factor in PE-2's extraordinary ability to thrive in the stressful environment found in large industrial fermentation tanks. In general, PE-2 persists during the whole season in a high concentration comparing to contaminant strains. Moreover, it has desirable characteristics, as high ethanol yield, low foam formation and low flocculation [1].

Flocculation is a cell-cell adhesion phenotype controlled by a variety of genetic mechanisms (FLO genes and their suppressors / activators) associated with medium characteristics, as cell density, carbon and/or nitrogen starvation, changes in pH, temperature, oxygen and agitation, type of sugar and levels of ethanol and cations [reviewed by 4; 5]. Flocculation is highly undesired during the industrial fermentation, since flocculent cells impair the centrifugation step and cell-substrate contact is reduced, increasing the fermentation time and resulting in loss of bioethanol yield [1,6].

At industrial scale, flocculation is also assigned as the co-aggregation of yeast and bacteria cells, which is often observed at high levels of bacterial contamination [6]. *Lactobacillus* species are the main contaminants found on sugarcane bioethanol production due to their ability to tolerate the stressful conditions during the fermentative process, such as high ethanol concentration [8 - 11% (v/v)], low pH imposed during anti-bacterial acid treatment (pH 2.0 – 3.0) and fermentation temperature (31 – 35°C) [7]. *L. plantarum* and, more prominently, *L. fermentum* have been related as the main biological factors responsible for the co-aggregation yeast under industrial scale [8].

The mannose-specific adhesin (Msa) identified in *L. plantarum* and *L. fermentum* has been assigned as responsible for the cell-cell interaction [9–11]. Recently, Hirayama et al.

(2012) constructed a collection of 12 deletion *S. cerevisiae* mutants with various incomplete mannan cell wall constituents structures. Among them, only *mnn2* mutant strain lost co-aggregation capacity due to *L. plantarum* action. Mnn2p is a mannosyltransferase that transfer the first α -1,2-linked mannose to mannan core structure to form the side chain that is extended after by Mnn5p [13]. Once Mnn2p is absent, the mutant constructed by Hirayama et al. (2012) has an unbranched chain mannan that prevent bacterium Msa linkage.

Despite *S. cerevisiae* is considered a vigorous and acid-tolerant organism [14], high concentrations of organic acids (lactic and acetic) produced by the bacterial contaminants associated to low pH and high concentration of ethanol reduces yeast metabolism and the fermentative fitness [15–17]. Narendranath et al. (2001) reported that the synergism among lactic and acetic acids induces an inhibitory effect on *S. cerevisiae* by reducing growth rate and decreasing rates of glucose consumption and ethanol production, when the concentrations of organic acids are present on medium at 0.5% (w/v) and 0.04% (w/v), respectively.

The genetic mechanisms that contribute to PE-2 adaptation during bioethanol production are of great importance to biotechnological purposes, however remain largely unknown [6,19]. Moreover, the stressful conditions found at industrial fermentation scale (biotic and abiotic) are distinct of the obtained by laboratories assays and not reproducible at bench scale [6]. Here, we investigate the transcriptional machinery of *S. cerevisiae* PE-2 strain under industrial co-aggregation phenotype due to bacterial contamination through genome-wide transcription profiling using high throughput sequencing-based methods (RNA-seq). Furthermore, we reported for the first time the physiology of PE-2 under industrial scale bioethanol fermentation under typical condition. Evidences that help to explain the superior PE-2 performance during industrial scale is also reported providing information to be used further on genetic improvements programs of bioethanol strains.

METHODS

Sample collection

Biological samples under two different industrial conditions were collected directly from bioethanol tanks from the same distillery. At the beginning of 2009 season, the PE-2 culture used as start inoculum became flocculent due to bacterial co-aggregation. To take advantage of this phenotype, we collected seven points of this fermentation (FL) along the process of one fed-batch cycle. Over the three months following, the cells were treated with antibiotics after each cycle and yeast community naturally reverted to the natural phenotype, and six points were

collected under typical fermentation (TF). Three biological replicates were collected for each one of the 13 time-points. After collection, samples were immediately transferred to a recipient containing dry ice for posterior kinetics analyses and RNA extraction.

Yeast genotyping

Unfrozen aliquots of each condition were plated in YPD solid medium (yeast extract 1% [w/v], peptone 2% [w/v], glucose 2% [w/v] and agar 2% [w/v]) for genotyping purpose. Colonies were isolated in YPD solid medium and DNA extraction was carried out by phenol-chloroform protocol [20]. Yeast strains were genotyped using four molecular markers previously designed based on the PE-2 genomic sequence (Carvalho-Netto *et al.*, data not published). PCR was performed using 20 colonies of each condition.

Bacteria isolation and 16S rDNA sequencing

Bacteria were isolated directly from bioethanol tanks and cultivated on LB solid medium (tryptone 1% [w/v], yeast extract 0.5% [w/v], NaCl 0.5% [w/v] and agar 2% [w/v]) in anaerobic conditions. In order to identify the bacteria species by 16S rDNA sequencing, colonies were grown on 5% sugar syrup medium. Bacteria DNA extraction protocol was adapted from Ausubel *et al.* (2002) using lysozyme (100 mg/mL) and proteinase K (10 mg/mL). The amplification of the 16S rDNA was performed by conducting a PCR in a final volume reaction of 50 μ L using 4 ng DNA, 0.5 μ M F27 (5' AGA GTT TGA TCM TGG CTC AG 3') and R1378 (5' CGG TGT GTA CAA GGC CCG GGA ACG 3') primers [21], 0.25 mM each dNTP, 3.5 mM $MgCl_2$, 1X Colorless GoTaq Flexi Buffer (Promega) and 1.25 U GoTaq Flexi DNA Polymerase (Promega). The amplification program consisted of one hold at 94°C for 3 min, followed by 40 cycles of 30 s at 94°C, 30 s at 55°C and 60 s at 72°C. A final 5-minutes extension was performed at 72°C. PCR products were purified using the same protocol described above for *RCE1* digestion.

PCR products (45 ng) were sequenced using the Big Dye Terminator kit (Applied Biosystems) on a 3500 Genetic Analyzer (Applied Biosystems). The reaction program consisted of one hold at 94°C for 2 min, followed by 35 cycles of 20 s at 94°C, 30 s at 55°C and 2 min at 60°C. Sequence similarity was obtained by BLASTn analysis using the GenBank non redundant (NR) Database (<http://blast.ncbi.nlm.nih.gov/>).

Fermentation kinetics

Aliquots of biological replicates were centrifuged and the supernatants were diluted in pure water (1:3). In order to determine the fermentation kinetics, diluted samples were filtered in Millipore 0.22 μm filters and analyzed by High Performance Liquid Chromatography (Alliance 2795 Waters) using a refractive index detector (HPLC-RI) and an Aminex HP-87H column (Bio-Rad). A standard curve was used to detect and quantify sucrose, glucose, fructose, ethanol, glycerol, acetic acid and lactic acid on samples.

RNA isolation and RNA-seq library preparation

Pools of biological material were obtained mixing the three biological replicates for each time-point. Total RNA of 13 fermentation time-points was extracted using phenol and chloroform following the protocols described by [20]. RNA-seq libraries were prepared from 1 μg total RNA following the manufacturer's protocol (Illumina). Briefly, mRNA was isolated using oligo(dt) magnetic beads and fragmented in the presence of divalent zinc ions. Fragmented RNA was then used for first and second strands cDNA synthesis. Double-stranded cDNA was end-repaired and 3' adenylated for ligation of sequence adapters. After ligation of adapters, fragments of approximately 250 bp were isolated by gel electrophoresis and PCR amplified. Libraries were validated on an Experion DNA chip (Bio-Rad) and quantified using a Qubit fluorometer (Invitrogen). Each RNA-seq library was sequenced in one lane of an Illumina Genome Analyzer IIx sequencer, which produced ~20-30 million reads (36 bp single-end).

Data analysis

For each RNA-seq library, reads were aligned against *S. cerevisiae* S288C genome (www.yeastgenome.org) using SOAPaligner version 2.20 (from short oligonucleotide alignment package) with the default settings. After that, a Perl script was made to calculate the number of reads aligned by genes from each RNA-seq library. Then the output file was analyzed by DEGSeq package for identifying differentially expressed (DE) genes (<http://www.bioconductor.org/packages/2.6/bioc/html/DEGseq.html>). Gene expression values were determined using RPKM formula [22]. Clustering and visualization of DE genes were obtained by the EXPANDER6 program [23].

Gene ontology (GO) terms of DE genes were obtained by SGD database (<http://www.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl>) using Yeast GO-Slim Process parameters and cutoff p-value <0.01. Functional GO enrichment terms were obtained using DE

genes between fermentations (TF1 vs. FL1; TF6 vs. FL7; TFs vs. FLs).

RNA-seq validation by Real Time qPCR (RT-qPCR)

In order to confirm RNA-seq data, 15 genes were analyzed by RT-qPCR in four pairwise comparisons (TF1 vs. TF4, TF1 vs. TF6, FL1 vs. FL4 and FL1 vs. FL7), totalizing 60 pairwise comparisons. List of genes and primers are presented as supplementary table 1. Aliquots of samples used to construct RNA-seq libraries were used in transcriptase reverse reaction to synthesize cDNA using *SuperScript Direct cDNA Labeling System* (Invitrogen), according to the procedures described by the manufacturer. The RT-qPCR mix consisted of 8 μ L SYBR Green Supermix (Bio-Rad, CA), 1 μ L each primer (0.5 μ M final concentration), 5 μ L water, and 1 μ L cDNA. The reaction program consisted of one hold at 95°C for 5 min, followed by 40 cycles of 15 s at 95°C and 75 s at 60°C. Fragments amplification and detection of *SYBR Green* (Applied Biosystems) were performed by *Step One Plus* thermocycler (Applied Biosystems). Relative expression ratio was calculated using the $2^{-\Delta\Delta CT}$ method [24]. *ACT1* and *YNL134c* genes were selected as endogenous genes to normalize expression values for TF and FL samples, respectively, since both genes showed little variation in expression in the different RNA-seq libraries.

***RCE1* partial amplification and enzymatic digestion**

The amplification of the *RCE1* partial sequence was performed by conducting a PCR in a final volume reaction of 50 μ L using 1 ng DNA, 0.5 μ M *RCE1_F* (5' ACC TTA TAT TGT GGA CCC GTT 3') and *RCE1_R* (5' CTC GAT AGA ATT CCA TAA TAG 3') primers, 0.25 mM each dNTP, 3.5 mM $MgCl_2$, 1X Colorless GoTaq Flexi Buffer and 1.25 U GoTaq Flexi DNA Polymerase (Promega). The amplification program consisted of one hold at 94°C for 2 min, followed by 35 cycles of 40 s at 94°C, 40 s at 56°C and 80 s at 72°C. A final 5-minutes extension was performed at 72°C. PCR products were purified using the NucleoSpin Extract II purification kit (Macherey-Nagel), according to the manufacturer's instructions. PCR products were digested using 10 U of *Mbol* (New England Biolabs Inc.), according to manufacturer's procedures. The patterns of digested bands were visualized by ethidium bromide staining of a 2% (w/v) agarose gel.

RESULTS AND DISCUSSION

RNA-seq reads alignment and microbial identification

Due to low microbiological control, the Brazilian bioethanol process involves a dynamic competition between the strain used as starter and a variety of wild yeast strains and bacteria that contaminate the fermentation [1,2]. At the beginning of 2009 season (April), it was observed at the industrial plant a significant flocculation phenotype driving on yeast culture. Once cell-cell adhesion was due to bacterial contamination (co-aggregation), use of antibiotic was coupled with acid treatment in order to reduce the contaminants and therefore the flocculation. This process took several days and the culture gradually back to the original non-flocculent (typical) condition. Despite the three months difference from flocculated and typical sampling, 95% of the yeast isolates were identified by genotyping analysis as PE-2 strain in both conditions, showing the high adaptation by this strain in such industrial plant.

Furthermore, the prevalence of PE-2 supports that gene expression data obtained is mainly derived from the strain of interest. The 13 RNA-seq libraries of TF and FL samples generated approximately 330 million reads (36 bases each), representing 11.88 Gb of sequence. Of those reads, approximately 76% aligned to reference genome and were assigned as PE-2 transcripts (Supplementary table 2).

RNA-seq libraries preparation procedure unavoidably carried a little variety of molecules that are not mRNA. Thereby, we took advantage of this shortcoming to search for sequences derivate from bacteria using reads that failed to align on the *S. cerevisiae* genome and were classified as ribosomal sequences. Reads assigned as bacteria were assembled into contigs and the similarity of the sequences was verified at GenBank database. In order to obtain a more reliable data, only long contigs (at least 1,400 bp) were used in the similarity search. Those contigs were classified exclusively as *Lactobacillus* in both libraries. Interestingly, TF and FL had distinct bacteria communities, comprising *L. helveticus* / *L. sobrius* and *L. fermentum*, respectively. The identities of bacteria previously isolated from flocculated samples were also confirmed by 16S rDNA sequencing. Finally, the isolates were remarkably able to flocculate (co-aggregate) a PE-2 non-flocculent culture when co-cultivated under laboratory conditions, similarly as the condition observed at the industrial fermentation.

Fermentation kinetics

Fermentation kinetics presented four main significant differences among fermentations compounds. In comparison to the typical fermentation, flocculated samples presented lower

bioethanol yield (i), lower glycerol content (ii), higher lactic and acetic acids concentrations (iii) and longest sucrose hydrolyzing rate (iv) (Figure 1).

Besides the longer time spent (FL: 21 hours; TF: 15 hours) and similar residual sugar (FL: 7.6g/L; TF: 4.5g/L), content of ethanol in FL was ~25% lower than in typical conditions, reaching 64.5g/L. Due to high time spent and low yield, we estimate that PE-2's flocculation led a decrease of approximately 15 million liters of bioethanol during the three months that the plant operated in such condition. As shown, flocculation is strictly detrimental and prohibitive during industrial process.

A remarkable aspect of PE-2 is the production of glycerol in low concentrations during bioethanol fermentation, once glycerol formation is inversely related to ethanol yield [1]. However, it is desirable that the strain produces enough glycerol to maintain the redox balance and the cellular osmoregulation [25,26]. Interestingly, TF samples had three times more glycerol concentration (4.74 g/L) in comparison to FL samples, suggesting that the FL cells were in metabolic misbalance and more susceptible to stressful condition imposed by industrial fermentation.

S. cerevisiae strains do not produce high amounts of organic acids (Abbott et al., 2009). Under laboratory conditions, PE-2 produces only 1.5 and 2.4 mg/L of lactic and acetic acid, respectively [15]. Therefore, organic acids generated during bioethanol production has been assigned to bacterial contaminants, mainly *Lactobacillus* species [7,28]. Here, at the last time point of fermentations, content of acetic and lactic acids was identified 6 and 3.5 times higher in FL than in typical fermentation, respectively (Figure 1). Since PE-2 produces low amounts of lactic and acetic acid [15] and the residual sugar contents were similar at the end of both fermentations, we speculate that sugars were consumed by bacterial contaminants and converted into organic acids.

Despite the lower content of ethanol identified in flocculated fermentation, glycolysis-related genes were in general not differentially expressed among conditions. Curiously, sucrose hydrolyzing gene, *SUC2*, was in general fourfold up-regulated in typical fermentation. Also, expression of *SUC2* at TF increased threefold when no more sugarcane juice was added to the tank and the level of C6 sugars (glucose and fructose) was reduce from 36 g/L to 6 g/L. This rapid activation of *SUC2* expression seems to be important for the prompt stress response of nutrient limitation (i.e., C6 sugars) during fermentation [29]. On the other hand, expression on FL decreased 7 times along the process. Such pattern could be partially explained by the presence of sucrose on FL along the whole fermentation (Figure 1) which provided a continuous supplying

of C6 sugars to the cell, leading the *SUC2* repression [30]. Moreover, the reduced yeast cells contact area with the medium due to cell-cell adhesion phenotype could also be associated with the distinct pattern observed on FL condition. Therefore, we can speculate that one of the reasons for the longer time spent at FL process is due to inefficient sucrose metabolism. If so, the shift of *SUC2* regulation expression by genetic engineering may be a potential strategy in such bacterial contamination condition.

Differential gene expression within fermentations

Time points TF1 and FL1 were used as reference to identify genes differentially expressed (DE) along time-points within fermentations. Using a p-value cutoff of 0.01, fold change > 2 (up-regulated) and fold change < -2 (down-regulated), we were able to identify a number of genes that are DE among different samples (Table 1).

In order to validate RNA-seq data, 15 genes were assessed by RT-qPCR totalizing 60 pairwise comparisons. The total expression trends of time-points analyzed were similar to 87% among techniques, with correlation values of R^2 : 0.7604 and R^2 : 0.7951, for FL and TF samples, respectively (Supplementary figure 1). Despite the absence of RNA-seq fermentation replicates, due to complexity of industrial process and transitory yeast flocculation phenotype, we obtained a high correlation values between techniques, supporting the majors transcriptional profiles obtained by our analysis.

Although *S. cerevisiae* is considered a vigorous and acid-tolerant organism [14], high concentrations of organic acids associated with low pH and high concentration of ethanol reduces its metabolism [15–17]. From the results presented on figure 2 it is possible to observe that the distribution of gene expression decreases when organic acids content reaches values above 4g/L on FL samples, suggesting a strong gene repression over the yeasts associated to content of organic acids (Figure 2). In this case, less genes were expressed and in lower levels in comparison to the previous time-points. At the low pH of fermentative conditions, the organic acids produced by contaminating bacteria will exist substantially in the undissociated state [31]. The undissociated organic acids present in the substrate diffuses across the cell membrane where will dissociate, generating protons that lower the intracellular pH and inhibit many metabolic functions [32]. Moreover, the dissociation produces anions that, being charged, may influence free radical production, leading to the severe oxidative stress [32]. Narendranath et al. (2001) reported that the synergism among lactic and acetic acids induces an inhibitory effect on *S. cerevisiae* by reducing growth rate and decreasing rates of glucose consumption and ethanol

production, when concentrations the organic acids are present on medium at 0.5% (w/v) and 0.04% (w/v), respectively.

Differential gene expression between fermentations

Gene expression comparisons between fermentations were performed using the time-points corresponding to the beginning (TF1 vs. FL1) and end of the processes (TF6 vs. FL7) in order to evaluate the transcriptional patterns of each condition. Moreover, we performed a global analysis using TF library versus FL library (TFs vs. FLs). Differentially expressed (DE) genes of global analysis were obtained using the gene expression averages of the six time-points of TF and the seven time-points of FL sample (Table 1C).

Besides the co-aggregative phenotype, genes related to flocculation (*MUC1*, *FLO5*, *FLO8*, *FLO9*, *FLO10* and *PHD1*) were not DE among fermentations, confirming that flocculation was not due to yeast genetic control. We observed once again that the main differences at transcriptional level among conditions were due to content variations of organic acids present on medium. The major plasma membrane H⁺-ATPase, encoded by *PMA1* [33], had no differential expression between samples at the beginning of fermentations (TF1 vs. FL1). However, we verified a two-fold *PMA1* induction in flocculated fermentation at the end of process. Pma1p related genes *AST1* (targeting factor to plasma membrane), *PMP1*, *PMP2* and *HRK1* (regulatory elements) had similar expression patterns (Figure 3B). These data shows that the mechanism to pump out protons in order to regulate cytoplasmic pH is more active on FL cells. This stress response, however, consumes excessive ATP and may cause an inhibition action by energy depletion [32].

The main transcriptional responses of *S. cerevisiae* in the presence of weak acids (lactic and acetic) are related to cell wall components, membrane-associate transport process and iron homeostasis [16,34,35]. It has been reported that Haa1p transcription factor and Haa1p-regulated genes are up-regulated in response to the lactic and acetic acids [35–37]. Among Haa1p target genes, *TPO2*, *YGP1*, *PHM8*, *GRE1*, *YPR157w*, *YER130c* and *HRK1* are also up-regulated in the FL7 sample in comparison to TF6 (Figure 3C). However, we did not observe differences on *HAA1* expression among fermentations, suggesting a co-regulation of those seven genes by distinct transcriptions factors [37].

During flocculated fermentation, cell wall related genes changed expression dramatically in comparison to typical fermentation. Those genes had a down-regulation expression varying from 2 - 17-fold along flocculated fermentation (Supplementary table 3). Kawahata et al. (2006)

reported that depletion in expression of the cell wall components *SED1*, *DSE2*, *CTS1*, *EGT2*, *SCW11*, *SUN4* and *TOS6* increased resistance to lactic acid in *S. cerevisiae*. Herein, PE-2 strain used the same mechanism in flocculated fermentation, down-regulating these seven genes at 3 - 6.8-fold in response to organic acids concentrations at the end of FL process (Figure 3D).

Gene ontology of DE genes

Gene ontology (GO) analyses were performed to identify functional signatures in gene expression using DE genes between fermentations (TF1 vs. FL1; TF6 vs. FL7; TFs vs. FLs). Two enriched GO terms presented meaningful patterns related to industrial fermentations: cellular amino acid and vitamin metabolic processes (Supplementary table 4).

On FL, several genes assigned as “cellular amino acid metabolic process” (*MET2*, *MET3*, *MET4*, *MET14*, *MET16*, *MET17*, *MET28*, *MET32*, *STR3* and *GSH1*) are involved mainly on methionine (MET) and glutathione (GHS) biosynthesis pathway (Figure 3E) GSH has an important role on *S. cerevisiae* protection against oxidative stress [38,39]. The first, and rate-limiting, step in the GSH biosynthetic pathway is catalyzed by the conjugation of glutamate and cysteine by *GSH1* (reviewed in 38). Once methionine is involved on cysteine biosynthesis, the expression profile of MET genes have an indirect effect on GSH biosynthesis by providing the supply of cysteine to the pathway [41]. Moreover, the transcription factors Met4p and Met32p, required for MET biosynthetic genes, are also essential for *GSH1* expression by cadmium-mediated regulation [42]. In addition of the gene repression observed previously, MET and GHS profiles suggest that yeasts on FL were in an oxidative stressful condition probably due to formation of intracellular reactive oxygen species triggered by lactic [43] and acetic acids [44].

On the other hand, vitamin metabolic process was identified prominently in TF samples. Interestingly, most of the genes (*PET18*, *PHO3*, *RPI1*, *THI2*, *THI3*, *THI4*, *THI13*, *THI20*, *SNO2* and *SNZ3*) participate on thiamine (vitamin B1) metabolic processes (Figure 3F). *SNO/SNZ* genes are involved in vitamin B1 and B6 biosynthesis and also have a role in oxidative stress tolerance [45–47]. Moreover, it was demonstrated at laboratory conditions that bioethanol strains carrying amplifications of these genes are less sensitive to fluctuations in levels of vitamin B1 presented in a medium with high sugar concentrations [48]; and it is suggested that those genes are important for an adaptive growth at industrial process [3,48]. Based on thiamine genes expression profile observed here and once *SNO2* and *SNZ3* are expressed under thiamine depletion [45], we conclude that sugarcane juice used during TF process should be deficient in

vitamin B1. Furthermore, we report by transcriptional analyses the relevance of thiamine genes associated to PE-2 strain adaptation during industrial Brazilian bioethanol production.

Differential allelic expression

PE-2 genome sequencing revealed that this strain is highly heterozygous, with a density of ~2 SNPs / kb [3]. Based on this, we took advantage of the high number of PE-2 sequences generated by RNA-seq reads (~9 Gb) to identify differences in allelic expression during fermentations and map possible events of loss of heterozygosity (LOH). Differential allelic expression (DAE) was obtained using an arbitrary threshold of 65%, i.e., more than 65% of the reads from a specific gene comes from one allele. When the coding region had more than one SNP, the DAE was assigned as the expression average of their alleles.

From the SNP detection analyses, 65 genes with DAE were identified exclusively on TF samples and 35 in FL (Supplementary table 5), suggesting a condition-dependent expression pattern regulated by specific transcriptional mechanisms. Furthermore, 214 genes were identified in both fermentations (Supplementary table 5). Interestingly, 144 genes located at the right arm of Chr13, from *FAR8* (YMR029c), had more than 95% of the reads derive from a unique allele (Figure 4A). This pattern suggests that the Chr13 region comprising approximately 600 kpbs might be homozygous on PE-2 industrial isolates (due to LOH events) or there is some genetic mechanism that drives toward this expression pattern.

In order to check the homozygous pattern on industrial isolates, we designed primers to amplify partially *RCE1* (YMR274c), which contains a SNP (A/G) at position 874 and affect the restriction cutting site of *Mbol* enzyme. *RCE1* sequences were amplified from PE-2 isolates from industrial fermentations using JAY270/PE-2 and FY23 strains as references. Digestion reactions confirmed the homozygosis for *RCE1* on PE-2 isolates from industrial plants and the opposite pattern on JAY270 (Figure 4B). Finally, using sequences from our PE-2 (JAY291) genome [3], we observed that the right arm of Chr13 is heterozygous between JAY291 and the reads from industrial PE-2 isolates. Taken together, we concluded that the isolated which has been propagated commercially lost the original heterozygosity on Chr13 right arm probably by crossing-over event.

For the remaining 70 genes with DAE identified in both fermentations, we supposed that the pattern observed here might be due to differences on *cis*-elements sequences, once such elements tend to be more condition-independent than *trans*-elements [49] and no DAE was detected among their direct flanking genes.

Conclusion

The gap in basic biological knowledge about PE-2 and related strains represents a significant barrier to their genetic improvement and the full exploitation of their biotechnological potential. It is consensus that genetic engineering of bioethanol strains should be preceded by genomic and transcriptomic studies to identify genetic characteristics associated to the yeast fermentative fitness [6,19]. The results presented here provide new insights into the biology of the PE-2 strain and allowed us the identification of stress response mechanisms during bioethanol production. Information derivate from industrial fermentations has great opportunities to be applied on PE-2 genetic improvement studies aiming an even superior fermentative fitness. At this point, the data presented here comprises a useful feature for such applications.

List of abbreviations

Single nucleotide polymorphism (SNP), differential expression (DE), differential allelic expression (DAE), chromosome (Chr), reverse transcriptase quantitative polymerase chain reaction (RT-qPCR), kilo base pairs (kbps), methionine (MET), glutathione (GHS), loss of heterozigosity (LOH).

Competing interests

The authors report no conflicts of interest.

Author's contributions

O.V.C.N. performed experiments, analyzed data and wrote the manuscript; M.F.C., L.S.M. and M.F.N. performed bioinformatics analysis; P.J.P.L.T. and P.A.M. prepared RNA-seq libraries and performed sequencing; L.A.L.C. performed kinetics analyses; J.L.A. performed allelic differential expression analysis and wrote the manuscript; G.A.G.P. designed the concept and wrote the manuscript.

All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Usina Nova América (Maracaí-SP) to provide biological material. This work was financially supported by ETH Bioenergia S/A and Brazilian funding agencies FAPESP (process number 2008/51500-5) and CNPq.

References

1. Basso LC, de Amorim HV, de Oliveira AJ, Lopes ML (2008) Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Research* 8: 1155–1163.
2. da Silva-Filho EA, Brito dos Santos SK, Resende ADM, de Moraes JOF, de Moraes MA, et al. (2005) Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting. *Antonie van Leeuwenhoek* 88: 13–23.
3. Argueso JL, Carazzolle MF, Mieczkowski P a, Duarte FM, Netto OVC, et al. (2009) Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Research* 19: 2258–2270.
4. Soares EV (2011) Flocculation in *Saccharomyces cerevisiae*: a review. *Journal of Applied Microbiology* 110: 1–18.
5. Verstrepen KJ, Klis FM (2006) Flocculation, adhesion and biofilm formation in yeasts. *Molecular Microbiology* 60: 5–15.
6. Amorim HV, Lopes ML, de Castro Oliveira JV, Buckeridge MS, Goldman GH (2011) Scientific challenges of bioethanol production in Brazil. *Applied Microbiology and Biotechnology* 91: 1267–1275.
7. Lucena BTL, dos Santos BM, Moreira JL, Moreira APB, Nunes AC, et al. (2010) Diversity of lactic acid bacteria of the bioethanol process. *BMC Microbiology* 10: 298.
8. Yokoya F, Oliva-Neto P (1991) Características da floculação de leveduras por *Lactobacillus fermentum*. *Brazilian Journal of Microbiology* 22: 12–16.
9. Pretzer G, Snel J, Molenaar D (2005) Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *Journal of Bacteriology* 187: 6128–6136.
10. Furukawa S, Nojima N, Nozaka S, Hirayama S, Satoh A, et al. (2012) Mutants of *Lactobacillus plantarum* ML11-11 Deficient in co-aggregation with yeast exhibited reduced activities of mixed-species biofilm formation. *Bioscience, Biotechnology, and Biochemistry* 76: 326–330.
11. Turner MS, Hafner LM, Walsh T, Giffard PM (2003) Peptide surface display and secretion using two LPXTG-containing surface proteins from *Lactobacillus fermentum* BR11. *Applied and Environmental Microbiology* 69: 5855–5863.
12. Hirayama S, Furukawa S, Ogihara H, Morinaga Y (2012) Yeast mannan structure necessary for co-aggregation with *Lactobacillus plantarum* ML11-11. *Biochemical and Biophysical Research Communications* 419: 652–655.

13. Rayner JC, Munro S (1998) Identification of the *MNN2* and *MNN5* mannosyltransferases required for forming and extending the mannose branches of the outer chain mannans of *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry* 273: 26836–26843.
14. Abbott DA, Zelle RM, Pronk JT, van Maris AJA (2009) Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. *FEMS Yeast Research* 9: 1123–1136.
15. Dorta C, Oliva-Neto P, -Abreu-Neto MS, Nicolau-Junior N, Nagashima AI (2005) Synergism among lactic acid, sulfite, pH and ethanol in alcoholic fermentation of *Saccharomyces cerevisiae* (PE-2 and M-26). *World Journal of Microbiology and Biotechnology* 22: 177–182.
16. Kawahata M, Masaki K, Fujii T, Iefuji H (2006) Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in *Saccharomyces cerevisiae* cultures induce expression of intracellular metal metabolism genes regulated by Aft1p. *FEMS Yeast Research* 6: 924–936.
17. Thomas KC, Hynes SH, Ingledew WM (2002) Influence of medium buffering capacity on inhibition of *Saccharomyces cerevisiae* growth by acetic and lactic acids. *Applied and Environmental Microbiology* 68: 1616–1623.
18. Narendranath NV, Thomas KC, Ingledew WM (2001) Effects of acetic acid and lactic acid on the growth of *Saccharomyces cerevisiae* in a minimal medium. *Journal of Industrial Microbiology & Biotechnology* 26: 171–177.
19. Argueso JL, Pereira GAG (2010) Perspective : Indigenous sugarcane yeast strains as ideal biological platforms for the delivery of next generation biorefining technologies. *International Sugar Journal* 112: 86–89.
20. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, et al. (2002) *Current Protocols in Molecular Biology*. 1.
21. Heuer H, Krsek M, Baker P, Smalla K, Wellington EM (1997) Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Applied and Environmental Microbiology* 63: 3233–3241.
22. Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5: 1–8.
23. Sharan R, Maron-Katz A, Shamir R (2003) CLICK and EXPANDER: a system for clustering and visualizing gene expression data. *Bioinformatics* 19: 1787–1799.
24. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C_T$) Method. *Methods San Diego Calif* 25: 402–408.

25. Bakker BM, Overkamp KM, van Maris AJ, Kötter P, Luttik M a, et al. (2001) Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. FEMS Microbiology Reviews 25: 15–37.
26. Nevoigt E, Stahl U (1997) Osmoregulation and glycerol metabolism in the yeast *Saccharomyces cerevisiae*. FEMS Microbiology Reviews 21: 231–241.
27. Ishida N, Saitoh S, Ohnishi T, Tokuhira K, Nagamori E, et al. (2006) Metabolic engineering of *Saccharomyces cerevisiae* for efficient production of pure L-(+)-lactic acid. Applied Biochemistry and Biotechnology 129-132: 795–807.
28. Oliva-Neto P, Yokoya F (1994) Evaluation of bacterial contamination in a fed-batch alcoholic fermentation process. World Journal of Microbiology and Biotechnology 10: 697–699.
29. Geng F, Laurent BC (2004) Roles of SWI/SNF and HATs throughout the dynamic transcription of a yeast glucose-repressible gene. The EMBO Journal 23: 127–137.
30. Ozcan S, Vallier LG, Flick JS, Carlson M, Johnston M (1997) Expression of the *SUC2* gene of *Saccharomyces cerevisiae* is induced by low levels of glucose. Yeast 13: 127–137.
31. Basso LC, Basso TO, Rocha SN (2010) Ethanol production in Brazil: The industrial process and its impact on yeast fermentation. In: Bernardes MA dos S, editor. Biofuel Production – Recent Developments and Prospects. Rijeka: InTech, Vol. 1530. p. 596.
32. Piper P, Ortiz Calderon C, Hatzixanthos K, Mollapour M (2001) Weak acid adaptation; the stress response that confers yeasts with resistance to organic acid food preservatives. Microbiology 147: 2635–2642.
33. Serrano R, Kielland-Brandt MC, Fink GR (1986) Yeast plasma membrane ATPase is essential for growth and has homology with (Na⁺ + K⁺), K⁺- and Ca²⁺-ATPases. Nature 319: 689–693.
34. Abbott DA, Knijnenburg TA, Poorter LMI, Reinders MJT, Pronk JT, et al. (2007) Generic and specific transcriptional responses to different weak organic acids in anaerobic chemostat cultures of *Saccharomyces cerevisiae*. FEMS Yeast Research 7: 819–833.
35. Abbott D a, Suij E, van Maris AJ a, Pronk JT (2008) Physiological and transcriptional responses to high concentrations of lactic acid in anaerobic chemostat cultures of *Saccharomyces cerevisiae*. Applied and Environmental Microbiology 74: 5759–5768.
36. Fernandes AR, Mira NP, Vargas RC, Canelhas I, Sá-Correia I (2005) *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. Biochemical and Biophysical Research Communications 337: 95–103.

37. Mira NP, Becker JD, Sá-Correia I (2010) Genomic expression program involving the Haa1p-regulon in *Saccharomyces cerevisiae* response to acetic acid. *Omics : a Journal of Integrative Biology* 14: 587–601.
38. Grant CM, MacIver FH, Dawes IW (1996) Glutathione is an essential metabolite required for resistance to oxidative stress in the yeast *Saccharomyces cerevisiae*. *Current Genetics* 29: 511–515.
39. Stephen DW, Jamieson DJ (1996) Glutathione is an important antioxidant molecule in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiology Letters* 141: 207–212.
40. Grant CM (2001) Role of the glutathione / glutaredoxin and thioredoxin systems in yeast growth and response to stress conditions. *Molecular Microbiology* 39.
41. Wheeler GL, Trotter EW, Dawes IW, Grant CM (2003) Coupling of the transcriptional regulation of glutathione biosynthesis to the availability of glutathione and methionine via the Met4 and Yap1 transcription factors. *The Journal of Biological Chemistry* 278: 49920–49928.
42. Dormer UH, Westwater J, McLaren NF, Kent NA, Mellor J, et al. (2000) Cadmium-inducible expression of the yeast *GSH1* gene requires a functional sulfur-amino acid regulatory network. *The Journal of Biological Chemistry* 275: 32611–32616.
43. Abbott DA, Sui E, Duong G-H, De Hulster E, Pronk JT, et al. (2009) Catalase overexpression reduces lactic acid-induced oxidative stress in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology* 75: 2320–2325.
44. Ludovico P, Rodrigues F, Almeida A, Silva MT, Barrientos A, et al. (2002) Cytochrome c release and mitochondria involvement in programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell* 13: 2598–2606.
45. Rodríguez-Navarro S, Llorente B, Rodríguez-Manzanque MT, Ramne A, Uber G, et al. (2002) Functional analysis of yeast gene families involved in metabolism of vitamins B1 and B6. *Yeast* 19: 1261–1276.
46. Padilla PA, Fuge EK, Crawford ME, Errett A, Werner-Washburne M (1998) The highly conserved, coregulated SNO and SNZ gene families in *Saccharomyces cerevisiae* respond to nutrient limitation. *Journal of Bacteriology* 180: 5718–5726.
47. Ehrenshaft M, Bilski P, Li MY, Chignell CF, Daub ME (1999) A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* 96: 9374–9378.
48. Stambuk BU, Dunn B, Alves SL, Duval EH, Sherlock G (2009) Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis. *Genome Research*: 2271–2278.

49. Tirosh I, Reikhav S, Levy AA, Barkai N (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* 324: 659–662.

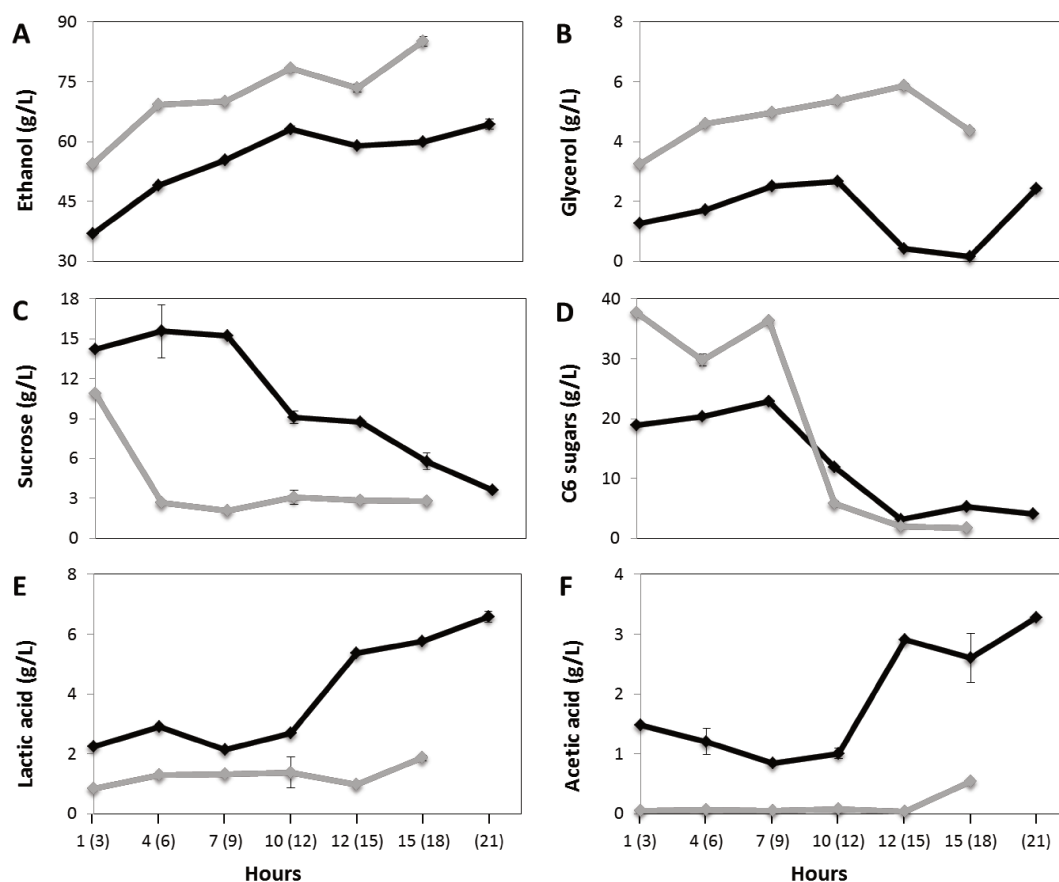


Figure 1. Kinetic of the major compounds identified during fermentations.

Content of ethanol (A); glycerol (B); C6 sugars, glucose and fructose, (C); sucrose (D); lactic acid (E); and acetic acid (F) were obtained by HPLC analyses using flocculated (black lines) and typical (gray lines) samples collected along fermentations. Hours regarding FL are presented in parentheses. Standard deviation bars were obtained using three techniques replicates for each time-point.

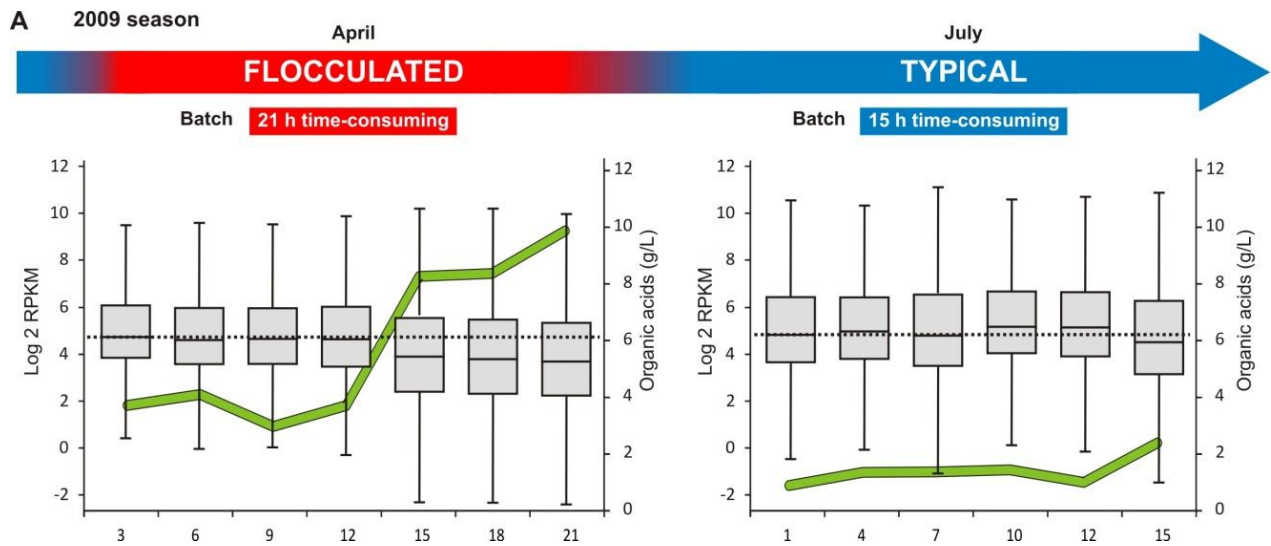


Figure 2. Gene expression distribution and a correlation with the content of organic acids.

A: Schematic representation of the fermentation conditions found at Nova América distillery along 2009 season. Boxplot of log2 RPKM for the (B) flocculated and the (C) typical conditions. Dot lines represent the median expression values of reference samples (FL1 and TF1). The concentration of organic acids (green lines) was obtained by the sum of lactic and acetic acids content identified for each sample.

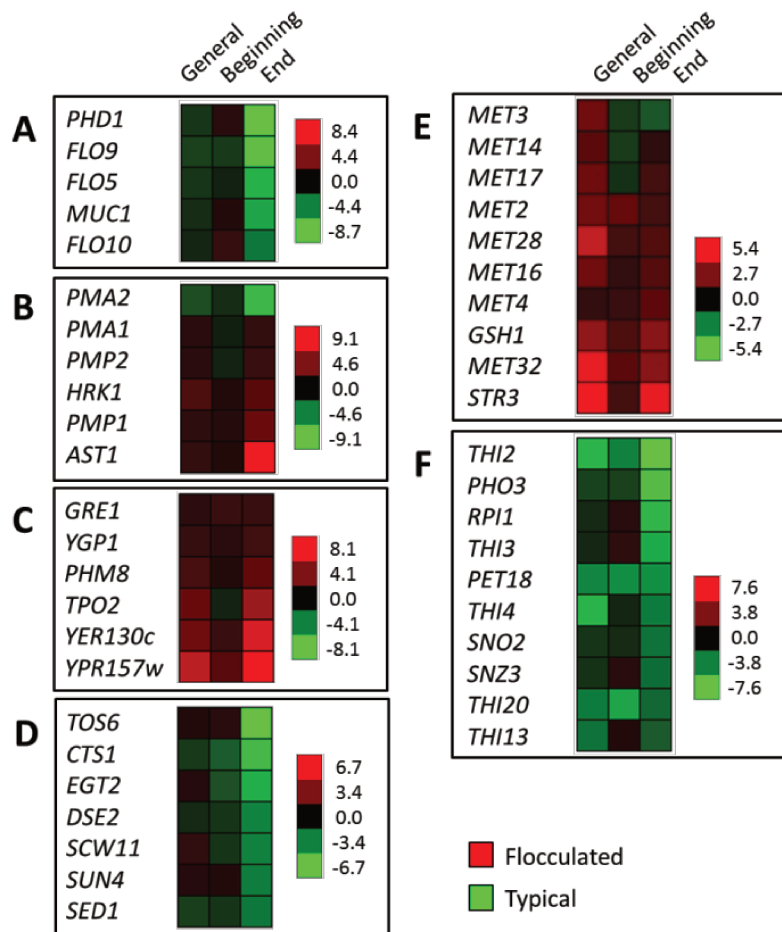


Figure 3. Gene expression comparison between typical and flocculated fermentations.

A- FLO genes and activators; B- Plasma membrane H^+ -ATPase (*PMA1*) and related genes; C- Haa1p target genes; D- Cell wall components; E- Methionine and glutathione related genes; F- Thiamine metabolic process genes. It was considered differentially expressed (DE) genes those with fold change ≥ 2 or ≤ -2 and p-value < 0.01 . Negative values are related to TF samples; positive values are related to FL samples. General (TFs vs. FLs) analysis was performed using six time-points of TF sample and seven time-points of FL sample. Beginning of fermentation is reported as TF1 vs. FL1 comparison; and end of fermentation is reported as TF6 vs. FL7 comparison. The software Expander6 was used for image draw using the end of fermentation as reference for gene clustering.

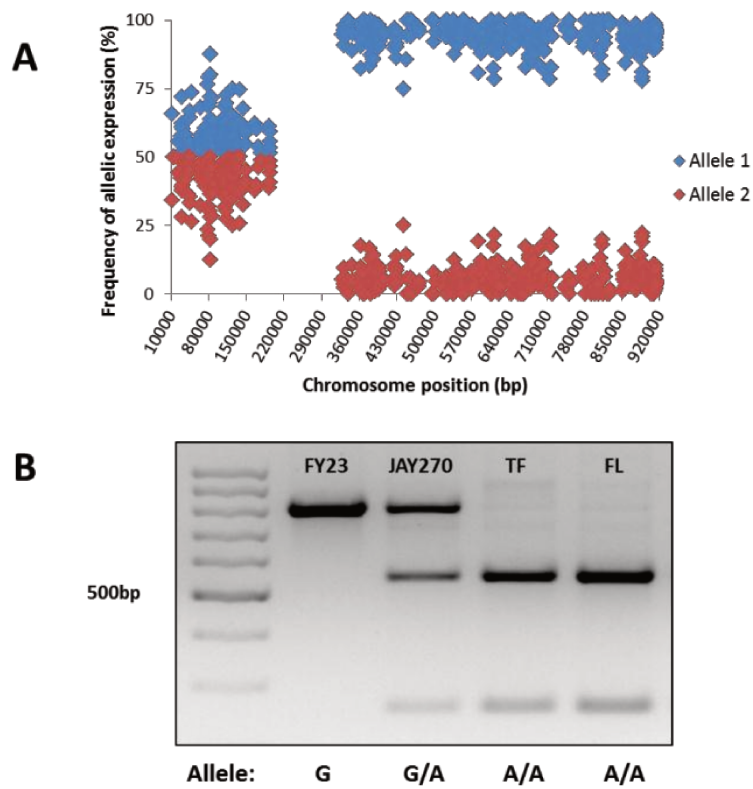


Figure 4. Frequency of allelic expression and pattern of enzymatic digestion.

A- The frequency of allelic expression of genes located at chromosome 13 (Chr13). B- Predict alleles of *RCE1* locus were determined by enzymatic digestion using *Mbol*: homozygous pattern, 543/256pb or 799bp; heterozygous pattern, 799/543/256bp. Molecular marker of 100bp was used at line 1.

Table 1. Number of differentially expressed (DE) genes during industrial bioethanol fermentation in two distinct conditions and a comparative analysis among them.

Assay			
A- Typical	DE genes	Down-regulated genes	Up-regulated genes
TF1 vs. TF2	211	134	77
TF1 vs. TF3	677	387	290
TF1 vs. TF4	1,263	359	904
TF1 vs. TF5	1,374	376	998
TF1 vs. TF6	1,624	1,004	620
B- Flocculated			
FL1 vs. FL2	240	161	79
FL1 vs. FL3	487	341	146
FL1 vs. FL4	1,461	990	471
FL1 vs. FL5	2,729	2,301	428
FL1 vs. FL6	2,797	2,402	395
FL1 vs. FL7	3,051	2,674	377
C- Comparative			
TF1 vs. FL1	398	63	335
TF6 vs. FL7	830	646	184
TFs vs. FLs	358	258	100

It was considered differential expressed genes those with expression ratio ≥ 2 or ≤ -2 and $p < 0.01$. Down-regulated genes refers to A) TF1; B) FL1; and C) TF1, TF6 and TFs. Up-regulated genes refers to A) TF2-TF6; B) FL2-FL7; C) FL1, FL7 and FLs.

CAPÍTULO 5

Desenvolvimento de linhagens de *Saccharomyces cerevisiae* geneticamente modificadas para produção de bioetanol

Osmar V. Carvalho-Netto¹, Juan Lucas Argueso², Katie Hummels³, Luís Carlos Basso⁴, Thalita P. Basso⁴, Gonçalo A. G. Pereira¹.

¹ Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas-SP, Brasil.

² Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins-CO, USA.

³ Department of Genetics, University of Iowa, Iowa City-IA, USA.

⁴ Departamento de Bioquímica, Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba-SP, Brasil.

Resumo

O processo de produção de bioetanol no Brasil é realizado principalmente por linhagens de *Saccharomyces cerevisiae* em um ambiente fermentativo não estéril, altamente competitivo e estressante. Estas leveduras devem, portanto, tolerar concentrações elevadas de etanol, persistirem ao longo de toda a safra e serem mais adaptadas ao processo que os contaminantes externos. As linhagens disponíveis comercialmente no Brasil são isolados selvagens selecionados diretamente das dornas por apresentarem excelentes características fermentativas. Entretanto, estas cepas vêm perdendo gradualmente seu desempenho ao longo dos anos, devido principalmente às alterações operacionais do processo. A manipulação genética visando o incremento de características de interesse industrial é uma alternativa promissora para o desenvolvimento de linhagens ainda mais robustas, uma vez que as cepas comerciais possuem um genótipo selvagem. Utilizando dados de transcriptoma gerados recentemente por nosso grupo, foram selecionados primeiramente sete promotores e dez genes para testar o conceito de modificação genética em uma das principais linhagens utilizadas atualmente, PE-2. Para cada um dos dez genes selecionados a região promotora original de apenas um cromossomo homólogo foi alterada de três maneiras diferentes. Dentre essas linhagens, nove se destacaram por apresentarem um perfil de competição superior em comparação com a linhagem parental. Estas linhagens foram então submetidas a testes fermentativos e novamente algumas tiveram um melhor desempenho em relação ao isolado selvagem, com um rendimento médio de até 0,4% superior durante oito ciclos de fermentação. Assim, nós acreditamos que os dados aqui apresentados referentes a estas duas características associadas, ainda que muito preliminarmente, abrem novas perspectivas para o desenvolvimento de uma nova geração de leveduras destinadas à produção de bioetanol.

Introdução

O processo fermentativo de produção de bioetanol no Brasil é realizado em condições não assépticas, o que promove a contaminação por bactérias e leveduras selvagens encontradas principalmente na cana-de-açúcar. Estes microrganismos competem pelo consumo de açúcares com as leveduras inoculadas no início da safra e reduzem o rendimento da fermentação, uma vez que nem todo o açúcar disponível é convertido em etanol ou é convertido em um tempo mais elevado (Amorim et al., 2011). Além disso, a maioria das leveduras contaminantes não possuem características desejáveis ao processo, como baixa floculação, alto rendimento e produtividade e persistência ao longo dos ciclos fermentativos (Basso et al., 2008).

Um fator mais agravante deste processo, em termos de produtividade, refere-se ao reciclo as quais as leveduras são submetidas ao longo de toda a safra, que compreende em média nove meses. Ao final de cada ciclo fermentativo (8 a 15 horas), todo o conteúdo das dornas é centrifugado para separação do vinho (mosto fermentado) e leveduras. Estas sofrem um tratamento ácido para redução da carga bacteriana e retornam à dorna para um novo ciclo fermentativo. Assim, as leveduras mais adaptadas ao processo fermentativo de determinada usina se sobressaem sobre as outras linhagens ao longo dos ciclos e dominam a população do processo. Dessa maneira, o prolongado sistema de reciclos de células favorece a formação de um ambiente altamente competitivo.

Aliado a isto, o processo de fermentação como um todo é altamente estressante em termos abióticos, exigindo das leveduras um elevado desempenho fermentativo (Amorim et al., 2011). Estes fatores tornam a escolha da levedura a ser inoculada no início da safra uma importante etapa na produção de bioetanol. Ao longo dos anos, foi sendo observado na maioria das usinas que as linhagens inoculadas eram rapidamente substituídas por linhagens selvagens (Basso et al., 2008; da Silva-Filho et al., 2005). Dessa forma, as linhagens com alta população ao final da safra passaram a ser isoladas e um extensivo trabalho foi realizado visando à identificação das características fisiológicas de cada isolado (Basso et al., 2008). Este e outros trabalhos relacionados resultaram na identificação de quatro principais cepas de *S. cerevisiae* (BG-1, CAT-1, SA-1 e PE-2) utilizadas atualmente. Estas linhagens passaram a ser utilizadas como inóculos comerciais nas usinas trazendo grandes benefícios para o setor, como a melhoria do rendimento e maior estabilidade do processo fermentativo (Basso et al., 2008).

Entretanto, ao longo das últimas duas décadas, estas linhagens vêm gradualmente

perdendo o desempenho observado anteriormente. Atualmente, estas linhagens permanecem em média em 46% das usinas onde foram implantadas e com uma população total de 52% (Luíz Carlos Basso, comunicação pessoal). Ao longo dos últimos 15 anos, o processo Brasileiro sofreu importantes alterações operacionais, como a redução da queima do canavial em substituição a colheita mecanizada e a condução do processo com teores de etanol mais elevados. Estes fatores podem ter contribuído para o surgimento de linhagens selvagens mais robustas em detrimento do desempenho das linhagens comerciais.

Diferentes processos que utilizam *S. cerevisiae* vêm incorporando características biotecnológicas de interesse industrial em suas linhagens através de manipulações genéticas pontuais (Donalies et al., 2008). Estas características são obtidas por diferentes estratégias, como a inserção e deleção de genes, mutações pontuais e a alteração da expressão gênica, entre outros. A aplicação desses organismos geneticamente modificados é principalmente destinada à panificação (Randez-Gil and Aguilera, 2003), produção de cerveja (Donalies and Stahl, 2002; Saerens et al., 2010), vinho (Cebollero et al., 2007; Cordier et al., 2007; Schuller and Casal, 2005) e bioetanol celulósico (Ha et al., 2011; Kuyper et al., 2004; Maris et al., 2007). Entretanto, as linhagens comerciais Brasileiras utilizadas na produção de bioetanol não se beneficiaram de tais tecnologias. Assim, é consenso que as linhagens comerciais Brasileiras têm potencial de se tornar ainda melhores através de modificações genéticas pontuais, uma vez que possuem um genótipo selvagem (Amorim et al., 2011; Argueso and Pereira, 2010).

Apesar do detalhado conhecimento fisiológico, apenas recentemente a estrutura e os mecanismos genéticos de regulação destas cepas vêm sendo estudados mais profundamente (Argueso et al., 2009; Babrzadeh et al., 2012; Carvalho-Netto et al., 2012 - dados não publicados; Stambuk et al., 2009). Os genomas de PE-2 e CAT-1 (Argueso et al., 2009; Babrzadeh et al., 2012), por exemplo, forneceram diversas informações sobre o padrão de heterozigosidade, rearranjos cromossômicos e a composição genética dessas importantes linhagens. Outro recente estudo, de análise global do transcriptoma de PE-2 sob condições industriais, revelou os mecanismos de utilização da maquinaria de transcrição da linhagem ao longo do processo fermentativo (Carvalho-Netto et al., 2012, dados não publicados). Além disso, o padrão de expressão para uma diversidade de genes pôde ser relacionado com a concentração dos compostos presentes no meio (etanol, açúcares e ácidos orgânicos). Neste caso, genes com variações de expressão ao longo da fermentação poderiam estar diretamente relacionados com o processo fermentativo, participando da cinética ou protegendo a célula

contra os estresses gerados pelos metabólitos do mesmo.

Baseado nestas informações, dez genes foram selecionados para serem utilizados em ensaios de alteração da expressão gênica na linhagem PE-2. As modificações, de modo a expressar os genes ao longo de toda a fermentação e em níveis mais elevados nestes mutantes, permitiram que as novas linhagens apresentassem um desempenho fermentativo superior em comparação com o isolado selvagem. Além disso, foi observada também uma maior robustez nos ensaios de competição, inferindo que estas linhagens poderiam competir com as linhagens selvagens do processo industrial de maneira mais eficiente. Dessa maneira, este trabalho apresenta o primeiro conjunto de linhagens geneticamente modificadas derivadas de PE-2 obtidas a partir de dados de transcrição global. Esta estratégia se mostrou eficiente e será aprimorada em estudos posteriores objetivando o desenvolvimento de linhagens altamente robustas para diferentes processos biotecnológicos.

MATERIAIS E MÉTODOS

Seleção dos genes de interesse

Os dados de expressão gênica utilizados neste estudo foram obtidos da linhagem de *S. cerevisiae* (PE-2) durante uma fermentação em escala industrial para produção de bioetanol (Carvalho-Netto et al., 2012, dados não publicados). Os genes e promotores utilizados neste estudo serão nomeados em forma de abreviações, uma vez que é intenção dos autores produzir uma publicação científica e proteger as possíveis tecnologias desenvolvidas sob forma de patente.

Os perfis de expressão gênica entre os pontos da fermentação industrial foram obtidos através de uma análise de agrupamento (<http://ep.ebi.ac.uk/EP/EPCLUST>) visando à identificação de genes com tendência similar de expressão entre as condições e os que apresentaram tendências distintas. Dessa forma, foram identificados genes/promotores cuja expressão é ativada ou reprimida no início, meio e ao final do processo de conversão do açúcar a etanol e quais são utilizados ao longo de todo o processo fermentativo.

Como critério de seleção, foram utilizados genes de funções desconhecidas ou com poucos estudos reportados, de maneira a gerar novos dados fenotípicos associados a estes genes. Para isto, os genes foram investigados quanto às suas funções e a participação em vias metabólicas utilizando os bancos de dados do KEGG (<http://www.genome.jp/kegg>) e SGD

(<http://www.yeastgenome.org/>). Além disso, utilizou-se também um segundo critério, que compreendeu a busca de genes com perfis de aumento da expressão ao longo do processo fermentativo.

Seleção dos promotores, construção dos plasmídeos e alteração da região promotora

A escolha das regiões promotoras foi baseada na avaliação da expressão gênica dos mesmos dados utilizados para a seleção dos genes descritos anteriormente. Foram utilizados quatro critérios nesta nova seleção: i- expressão gênica com pouca variação ao longo da fermentação; ii- diferentes níveis de expressão entre os promotores selecionados; iii- região promotora de tamanho não superior a 1.100 pares de bases; iv- região promotora referente a um único gene.

Para substituição da região promotora original dos genes selecionados, foi utilizado o plasmídeo pFA6KanMX4. Este plasmídeo contém o gene KanMX4 que confere resistência ao antibiótico geneticina, permitindo dessa maneira a seleção das células transformadas (Figura 1A). Para a primeira etapa de construção do plasmídeo, a região promotora selecionada foi amplificada por PCR a partir do DNA do isolado JAY291 (haplóide de PE-2). Aos iniciadores sintetizados para amplificação da região de interesse, foram adicionados 12 bases específicas com o objetivo de criar um sítio de corte para a enzima *EcoRI* e outro para *SacII*. O esquema da região promotora de T1 (pT1) amplificada após PCR é apresentado na figura 1B. Após amplificação da região promotora, o fragmento foi digerido com as enzimas de restrição *EcoRI* e *SacII*. O mesmo procedimento foi realizado para o plasmídeo pFA6KanMX4, que contém apenas um sítio de corte para cada enzima. Com ambos os fragmentos digeridos, estes foram ligados através de uma reação contendo a enzima ligase (Figura 1C).

A alteração da região promotora foi obtida via recombinação homóloga após transformação das leveduras. Para cada gene em estudo foram utilizados três diferentes construções de plasmídeos, objetivando alterar a expressão do gene de maneiras diferenciadas. Na figura 2 é apresentada esquematicamente a metodologia utilizada para alteração da região promotora do gene H2.

Transformação de leveduras

Para a transformação da linhagem PE-2 com o DNA de interesse foi utilizado o isolado

diplóide JAY270. Para o preparo das leveduras competentes, as células foram inoculadas em 5 mL de meio líquido YEPD (2% de glicose) e incubadas sob agitação por 30°C durante 16 horas. Esta cultura foi então transferida para 200 mL de meio líquido YEPD e mantidas a 30°C por 4 horas sob agitação (200 rpm) para crescimento exponencial. Após este período, o meio foi centrifugado por 5 minutos a 5000 g, descartado o sobrenadante e as células ressuspensas em 20 mL de água estéril. As células foram novamente centrifugadas e o sobrenadante foi descartado. Às células foi adicionado 1 mL de solução 1X TE/LiAc e estas foram ressuspensas.

Em um tubo de 1,5 mL, o DNA de interesse (12,5 µL) foi misturado com 100 µg de esperma de salmão previamente denaturado (10 minutos a 99°C). Cem µL de células competentes foram adicionados e misturados com a solução de DNA. A esta solução, foram adicionados 600 µL de 40% PEG/1X LiAc (800 µL 50% PEG; 100 µL 10X LiAc; 100 µL água). Após homogeneização, as células foram incubadas a 30°C por 30 minutos sob agitação (200 rpm). Em seguida, 70 µL de DMSO foram adicionados e misturados por inversão.

As células foram então aquecidas a 42°C por 15 minutos em banho-maria. Após este período, as células foram imediatamente transferidas para o gelo e mantidas por 2 minutos. Foram adicionados 700 µL de água, a solução foi misturada por inversão e centrifugada por 30 segundos a 13000 rpm. O sobrenadante foi descartado, as células foram ressuspensas com a água restante no tubo, plaqueadas em meio sólido YEPD e incubadas a 30°C. Após 18 horas, foram feitas réplicas em placas contendo o antibiótico de interesse. A cultura foi então incubada a 30°C por 2 a 4 dias até a formação de colônias.

Confirmação da alteração da região promotora

Todas as linhagens com a região promotora original modificada foram analisadas quanto à conservação da região codante e a exata localização da sequência inserida. Para isto foram amplificadas por PCR a região que compreende parte do gene de resistência ao antibiótico (KanMX4) à região interna do gene. Após a confirmação do tamanho da sequência amplificada, o produto foi sequenciado e alinhado com a sequência codante original.

Ensaio de competição

Estes ensaios compreendem o crescimento de duas linhagens diferentes em um mesmo meio para testar a adaptação das mesmas sob as condições analisadas. Nestas competições

foram utilizadas as linhagens com a região promotora alterada, que possuem o gene KanMX4, em competição com o isolado JAY270 derivado da linhagem PE-2 e sem nenhuma marca de resistência a antibiótico.

Inicialmente, as linhagens com a região promotora alterada foram individualmente misturadas com a mesma quantidade de células de JAY270. Alíquotas de 25 µL de cultura (diluição 10^{-4}) foram plaqueadas em ambos meios sólidos de YEPD contendo o antibiótico geneticina ou sem antibiótico, para verificar a razão de células resistentes a geneticina (região promotora alterada) em relação ao número total de células. Assim, uma razão de 1 Kan : 2 YEPD significa o mesmo número de células das duas linhagens no meio de fermentação.

As culturas foram então inoculadas em 5 mL de xarope de cana-de-açúcar com 18% de sacarose e mantidas a 30°C por 24 horas para permitir o crescimento até a saturação de células e a máxima conversão de açúcar a etanol. Ao final da incubação, uma alíquota de 25 µL de cultura (0,5% de inóculo) foi transferida para novo meio de fermentação. Este processo foi repetido durante 10 dias consecutivos. Nos ciclos cinco e dez, 25 µL de cultura (diluição 10^{-4}) foram novamente plaqueados em ambos meios sólidos de YEPD para verificar a razão de células resistentes a geneticina em relação ao número total de células.

Desvios progressivos da razão inicial 1: 2 foram observados quando uma linhagem estava crescendo mais vigorosamente que outra.

Ensaio fermentativo e quantificação de compostos

O isolado JAY270 foi usado como referência para determinar o desempenho fermentativo das linhagens modificadas utilizando o protocolo de Basso e colaboradores (2008). As leveduras foram propagadas a 30°C em melaço de cana-de-açúcar contendo 10% de açúcares totais (m/v) em estufa sem agitação. O volume do meio foi dobrado a cada 24 horas e a biomassa celular foi coletada após centrifugação quando alcançou a massa desejada. As fermentações foram conduzidas a 33°C em meio completo (mosto) contendo 20% de açúcares totais (composto por 50% de caldo de cana-de-açúcar e 50% de melaço). A biomassa úmida de leveduras foi adicionada de maneira a apresentar uma concentração final de 15% do volume total da fermentação, simulando as condições industriais. O mosto foi adicionado em três porções iguais a cada 1,5 hora após o inóculo das células. Ao final de cada ciclo fermentativo, as células foram coletadas por centrifugação (800g, 20 minutos), pesadas, tratadas com ácido sulfúrico (pH 2,5 por 1 hora) e reutilizadas em um novo processo, totalizando oito ciclos fermentativos.

Os compostos das duplicatas de fermentação foram analisados quanto ao conteúdo de etanol (após destilação seguida de quantificação em densímetro eletrônico AP Paar), glicerol, sacarose, glicose e frutose (HPAEC, Dionex DX-300). A viabilidade celular foi obtida por coloração com azul de metileno e contagem de células em microscópio ótico.

RESULTADOS

Seleção de genes e promotores e construção dos plasmídeos

Para avaliar a estratégia de identificação de fatores importantes para o desempenho fermentativo da linhagem, dez genes foram selecionados baseado nos critérios mencionados anteriormente. Estes genes apresentaram expressões ao longo da fermentação variando entre 2,2 a 96 vezes. Os valores de expressão gênica e a descrição resumida das funções dos genes selecionados são apresentados na tabela 1.

Devido às características apresentadas, nós hipotetizamos que estes genes poderiam estar relacionados de alguma maneira ao processo fermentativo participando diretamente da fermentação ou protegendo a célula contra os estresses do processo, principalmente concentração de etanol e estresse oxidativo. Desta maneira, modificações genéticas pontuais visando um aumento na expressão do gene e uma ativação prévia antes dos estresses serem detectados em níveis suficientes para ativação da expressão, poderiam proteger a célula com uma maior eficiência e aumentar o desempenho fermentativo.

Para testar esta hipótese, a região promotora desses dez genes foi substituída de maneira a torná-los constantemente expressos durante toda a fermentação e em níveis superiores aos originais. Para isto, sete genes foram selecionados para obtenção de suas regiões promotoras e utilização na construção de plasmídeos contendo o gene KanMX4 como repórter para identificação das células transformadas (Tabela 2).

Linhagens geneticamente modificadas

Baseado nos valores de expressão apresentados nas tabelas 1 e 2, foram selecionados três promotores para serem utilizados com cada um dos dez genes de interesse. O objetivo da utilização de três promotores individualmente foi obter transformantes com uma expressão constante ao longo da fermentação e em três níveis diferentes. Assim, o maior número de alterações na expressão de um único gene aumenta a possibilidade de escolha das melhores combinações para prosseguimento dos ensaios com as linhagens modificadas. As

combinações dos plasmídeos utilizados contendo os promotores assim como o aumento da expressão esperada para cada modificação são apresentadas na tabela 3. A confirmação da alteração de expressão das linhagens mutantes será realizada em um segundo momento utilizando apenas as linhagens que apresentarem alguma característica de interesse.

No total, foram construídas 29 linhagens de PE-2 com a região promotora alterada. Apenas uma alteração de promotor não foi obtida, utilizando o gene Y4w em combinação com a região promotora pA1 (Tabela 4).

Ensaio de competição

Em um primeiro ensaio de competição para verificar o efeito das modificações nas 29 linhagens obtidas pela substituição da região promotora, foi utilizado como controle o isolado JAY270 sem nenhuma marca de resistência a antibiótico. Dessa maneira, a razão de células presentes no meio de fermentação foi obtida pela razão do número de colônias presentes na placa com geneticina (linhagens com a região promotora modificada) e das colônias presentes na placa de YEPD sem antibiótico. Assim, uma razão de 1:2 (Kan e YEPD, respectivamente) significa que as duas linhagens estão presentes no meio com o mesmo número de células.

Desvios nas proporções originais dos ensaios (aproximadamente 1:2) puderam ser observados em alguns casos para as diferentes competições. Ao final dos 10 ciclos de fermentação, o controle utilizado no ensaio (JAY270 vs. JAY270+Kan) manteve uma razão de células próxima do constante (Figura 3A). Entretanto, 16 das 29 linhagens apresentaram um efeito neutro de crescimento (sem grandes alterações na razão), quatro linhagens apresentaram redução na razão do número de células com o promotor alterado e nove tiveram um efeito positivo, sugerindo que estas últimas possuem um crescimento mais vigoroso em relação à linhagem parental (JAY270). Na figura 3 são apresentados os gráficos das razões do número de células para o ensaio controle e para os outros ensaios com dois exemplos de cada comportamento observado.

Ensaio fermentativos

As linhagens que apresentaram um perfil de crescimento positivo nos ensaios de competição foram avaliadas novamente para observação do rendimento fermentativo em comparação com a linhagem parental (JAY270). Dessa maneira, foram avaliadas as nove linhagens derivadas de PE-2 utilizando oito ciclos fermentativos. Com o intuito de aumentar o estresse imposto às leveduras, o teor alcoólico da fermentação foi sendo gradualmente elevado

de 8% inicialmente para 14% (v/v) ao final do último ciclo. A viabilidade celular ao final do último ciclo foi bastante elevada para todas as linhagens, com uma média de 99%.

Curiosamente, o rendimento de algumas linhagens modificadas foi superior nas fermentações com elevado grau alcoólico (sétimo e oitavo ciclos) em referência ao padrão JAY270 (Tabela 5). Dentre estas, destacam-se as linhagens OY26, OY28 e OY49, com um incremento médio variando entre 0,3% a 0,77%. Apesar de os incrementos serem relativamente baixos, o teste estatístico T mostrou que são significativamente diferentes, sugerindo que estas linhagens toleram melhor elevadas concentrações de etanol que o isolado selvagem. Além disso, estas linhagens apresentaram um rendimento superior médio ao longo dos oito ciclos fermentativos de até 0,36%, ou seja, um incremento de 3% considerando os açúcares não metabolizados pela linhagem referência (JAY270).

DISCUSSÃO

Neste trabalho, apresentamos dados relacionados à uma fase do ciclo de estudos para compreensão e manipulação da linhagem PE-2. Apesar de dados promissores, a manipulação dessas linhagens foi realizada em um isolado não domesticado e utilizando informações preliminares do transcriptoma com o objetivo de testar o conceito. Neste caso, experimentos complementares estão sendo desenvolvidos com o intuito de validar os dados aqui apresentados e gerar mais informações associadas aos fenótipos observados.

Os dados de expressão gênica que estão sendo gerados por nosso grupo e a variedade de informações do genoma de PE-2 ainda não exploradas são um grande potencial de interesse biotecnológico. Entre estas possibilidades, pode-se destacar a utilização de outras combinações de genes/promotores nos ensaios de manipulação genética, assim como a variação do tamanho do promotor de interesse e o número de *locus* que este controla. Além disso, dados de expressão alélica diferencial de PE-2 indicam que em determinados genes há uma preferência de expressão de um alelo pelo outro e que esta variação pode ser tanto dependente quanto independente da condição (Carvalho-Netto et al., 2012, dados não publicados). Neste caso, os elementos *cis* e *trans* podem ser modificados de maneira a se obter uma regulação mais controlada das características de interesse (Alper et al., 2005; Ellis et al., 2009). Finalmente, há ainda regiões genômicas que foram identificadas em PE-2, mas que encontram-se ausentes na linhagem padrão de laboratório (S288c) e em outras linhagens relacionadas (Argueso et al., 2009).

Este estudo demonstrou que novas linhagens de *S. cerevisiae* podem ser eficientemente desenvolvidas utilizando a base genética e as informações de genômica e transcriptoma das linhagens comerciais utilizadas no Brasil para produção de bioetanol. Acreditamos que estas linhagens possam ser aplicadas não apenas para produção deste biocombustível a partir de cana-de-açúcar como também para outros processos biotecnológicos, como, por exemplo, a produção de bioetanol celulósico.

Referências

- Alper H, Fischer C, Nevoigt E, Stephanopoulos G. 2005. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America* **102**:12678-83.
- Amorim HV, Lopes ML, de Castro Oliveira JV, Buckeridge MS, Goldman GH. 2011. Scientific challenges of bioethanol production in Brazil. *Applied Microbiology and Biotechnology* **91**:1267-75.
- Argueso JL, Pereira GAG. 2010. Perspective: Indigenous sugarcane yeast strains as ideal biological platforms for the delivery of next generation biorefining technologies. *International Sugar Journal* **112**:86-89.
- Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OVC, Missawa SK, Galzerani F, Costa GGL, Vidal RO, Noronha MF, Dominska M, Andrietta MGS, Andrietta SR, Cunha AF, Gomes LH, Tavares FCA, Alcarde AR, Dietrich FS, McCusker JH, Petes TD, Pereira GAG. 2009. Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Research* **19**:2258-70.
- Babrzadeh F, Jalili R, Wang C, Shokralla S, Pierce S, Robinson-Mosher A, Nyren P, Shafer RW, Basso LC, de Amorim HV, de Oliveira AJ, Davis RW, Ronaghi M, Gharizadeh B, Stambuk BU. 2012. Whole-genome sequencing of the efficient industrial fuel-ethanol fermentative *Saccharomyces cerevisiae* strain CAT-1. *Molecular Genetics and Genomics* **287**:485-95.
- Basso LC, de Amorim HV, de Oliveira AJ, Lopes ML. 2008. Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Research* **8**:1155-63.
- Cebollero E, Gonzalez-Ramos D, Tabera L, Gonzalez R. 2007. Transgenic wine yeast technology comes of age: is it time for transgenic wine? *Biotechnology Letters* **29**:191-200.
- Cordier H, Mendes F, Vasconcelos I, François JM. 2007. A metabolic and genomic study of engineered *Saccharomyces cerevisiae* strains for high glycerol production. *Metabolic Engineering* **9**:364-378.

- Donalies UEB, Nguyen HTT, Stahl U, Nevoigt E. 2008. Improvement of *Saccharomyces* yeast strains used in brewing , wine making and baking. *Yeast* **111**:67-98.
- Donalies UEB, Stahl U. 2002. Increasing sulphite formation in *Saccharomyces cerevisiae* by overexpression of *MET14* and *SSU1*. *Yeast* **19**:475-484.
- Ellis T, Wang X, Collins JJ. 2009. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnology* **27**:465-471.
- Ha S-J, Galazka JM, Kim SR, Choi J-H, Yang X, Seo J-H, Glass NL, Cate JHD, Jin Y-S. 2011. Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. *Proceedings of the National Academy of Sciences of the United States of America* **108**:504-9.
- Kuyper M, Winkler AA, van Dijken JP, Pronk JT. 2004. Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle. *FEMS Yeast Research* **4**:655-64.
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**:402-408.
- Maris AJ, Winkler AA, Kuyper M, Laat WT, Dijken JP, Pronk JT. 2007. Development of efficient xylose fermentation in *Saccharomyces cerevisiae*: xylose isomerase as a key component. *Advances in Biochemical Engineering Biotechnology* **108**:179-204.
- Randez-Gil F, Aguilera J. 2003. Baker's yeast: challenges and future prospects. *Topics in Current Genetics* **2**:57-98.
- Saerens SMG, Duong CT, Nevoigt E. 2010. Genetic improvement of brewer's yeast: current state, perspectives and limits. *Applied Microbiology and Biotechnology* **86**:1195-1212.
- Schuller D, Casal M. 2005. The use of genetically modified *Saccharomyces cerevisiae* strains in the wine industry. *Applied Microbiology and Biotechnology* **68**:292-304.
- da Silva-Filho EA, Brito dos Santos SK, Resende ADM, de Morais JOF, de Morais MA, Ardaillon Simões D. 2005. Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting. *Antonie van Leeuwenhoek* **88**:13-23.
- Stambuk BU, Dunn B, Alves SL, Duval EH, Sherlock G. 2009. Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis. *Genome Research*:2271-2278.

Tabela 1. Valores de expressão gênica e descrição resumida da função de dez genes selecionados para estudos de manipulação genética.

Genes	RPKM						Variação de expressão (X)	Descrição
	1h	4h	7h	10h	12h	15h		
R3	242	288	316	529	1275	3748	15,5	Proteína de função desconhecida
S1	21	51	62	530	642	2023	96	Proteína requerida para tolerância a altas temperaturas
D2	151	522	798	4999	5088	6345	42	Proteína de resposta a fatores de estresses
H2	582	515	202	640	814	2053	3,5	Proteína <i>heat shock</i> com atividade chaperona
M1	443	322	217	407	602	965	2,2	Co-ativador transcricional
Y4c	384	305	516	1174	1191	1246	3,2	Proteína putativa de função desconhecida
T2	27	83	188	485	679	883	32,7	Proteína induzida por estresse
H1	117	145	192	296	328	453	3,9	Proteína receptora e transdutora de níveis de hidroperóxido
Y6	482	479	330	539	1461	8091	16,8	Proteína de função desconhecida
Y4w	464	672	904	1637	2618	7768	16,7	Proteína putativa de função desconhecida

RPKM: valores de expressão gênica em *Reads Per Kilobase of exon model per Million mapped reads*. Amostras coletadas em condições típicas de usina, em seis pontos ao longo do processo, correspondendo ao início (1 hora) e final da fermentação (15 horas). A descrição dos genes foi obtida no banco de dados do SGD (<http://www.yeastgenome.org/>).

Tabela 2. Promotores utilizados nos ensaios de manipulação genética da linhagem PE-2.

Promotor	ID	RPKM (média ± DP)	Tamanho (pb)
pI2	OVP1	465 ± 113	302
pP1	OVP2	730 ± 116	297
pC4	OVP3	1642 ± 267	329
pT1	OVP4	6170 ± 2297	591
pF1	OVP5	10790 ± 2471	639
pA1	OVP6	13066 ± 2893	1081
pE1	OVP7	26185 ± 12379	1065

ID: identificação do plasmídeo contendo o promotor. RPKM: média dos valores de expressão gênica em *Reads Per Kilobase of exon model per Million mapped reads* baseado em seis pontos ao longo da fermentação em escala industrial. DP: desvio padrão. Pb: pares de bases.

Tabela 3. Combinação dos promotores utilizados nos ensaios de manipulação gênica para os dez genes selecionados.

Genes	Construções utilizadas (OVP)			Aumento de expressão estimada (vezes)*		
	A	B	C	A	B	C
R3	4	6	7	1,6	3,5	7
S1	3	4	6	1,2	3	6,5
D2	4	6	7	1	2	4
H2	3	4	6	1,2	3	6,5
M1	3	4	5	1,7	6,3	11
Y4c	3	4	6	1,3	5	10,5
T2	2	3	4	1,2	1,9	7
H1	1	3	4	1	3,5	13
Y6	5	6	7	1,3	1,6	3
Y4w	5	6	7	1,4	1,7	3,4

Foram realizadas três alterações (A, B e C) na região dos promotores originais para cada gene alvo utilizando como molde os sete plasmídeos construídos anteriormente (OVP1 a OVP7). *Aumento de expressão estimada ao longo da fermentação baseada no valor de expressão do gene ao qual o promotor foi obtido.

Tabela 4. Linhagens derivadas de PE-2 desenvolvidas nos ensaios de manipulação genética utilizando 10 genes selecionados.

Gene	Isolado	Genótipo	Gene	Isolado	Genótipo
R3	OVY17	JAY270 R3::KanMX4pT1/R3	Y4c	OVY52	JAY270 Y4c::KanMX4pC4/Y4c
	OVY18	JAY270 R3::KanMX4pA1/R3		OVY53	JAY270 Y4c::KanMX4pT1/Y4c
	OVY19	JAY270 R3::KanMX4pE1/R3		OVY54	JAY270 Y4c::KanMX4pA1/Y4c
S1	OVY20	JAY270 S1::KanMX4pC4/S1	T2	OVY55	JAY270 T2::KanMX4pP1/T2
	OVY21	JAY270 S1::KanMX4pT1/S1		OVY56	JAY270 T2::KanMX4pC4/T2
	OVY22	JAY270 S1::KanMX4pA1/S1		OVY57	JAY270 T2::KanMX4pT1/T2
D2	OVY23	JAY270 D2::KanMX4pT1/D2	H1	OVY58	JAY270 H1::KanMX4pI2/H1
	OVY24	JAY270 D2::KanMX4pA1/D2		OVY59	JAY270 H1::KanMX4pC4/H1
	OVY25	JAY270 D2::KanMX4pE1/D2		OVY60	JAY270 H1::KanMX4pT1/H1
H2	OVY26	JAY270 H2::KanMX4pC4/H2	Y6	OVY61	JAY270 Y6::KanMX4pF1/Y6
	OVY27	JAY270 H2::KanMX4pT1/H2		OVY62	JAY270 Y6::KanMX4pA1/Y6
	OVY28	JAY270 H2::KanMX4pA1/H2		OVY63	JAY270 Y6::KanMX4pE1/Y6
M1	OVY49	JAY270 M1::KanMX4pC4/M1	Y4w	OVY64	JAY270 Y4w::KanMX4pF1/Y4w
	OVY50	JAY270 M1::KanMX4pT1/M1		OVY65	JAY270 Y4w::KanMX4pE1/Y4w
	OVY51	JAY270 M1::KanMX4pF1/M1			

Tabela 5. Rendimento fermentativo das linhagens geneticamente modificadas em comparação com o isolado selvagem.

Ciclo	Teor alcoólico médio (% v/v)	Rendimento (%)									
		OVY17	OVY19	OVY21	OVY23	OVY24	OVY26	OVY28	OVY49	OVY64	JAY270
1	7,93	89,34	89,41*	89,79*	87,59	89,52	89,31*	90,21*	89,93*	89,34	89,09
		(+ - 0,18)	(+ - 0,09)	(+ - 0,22)	(+ - 1,97)	(+ - 0,39)	(+ - 0,04)	(+ - 0,10)	(+ - 0,37)	(+ - 0,32)	(+ - 0,17)
2	8,82	89,97	90,77*	89,87	90,85	90,88*	90,54	90,42	91,39*	90,35	90,43
		(+ - 0,36)	(+ - 0,00)	(+ - 0,18)	(+ - 0,53)	(+ - 0,10)	(+ - 0,40)	(+ - 1,06)	(+ - 0,31)	(+ - 0,03)	(+ - 0,06)
3	8,80	88,46	89,52	88,67	89,34	88,90	89,06	89,59	89,73	89,77	89,84
		(+ - 0,99)	(+ - 0,11)	(+ - 0,41)	(+ - 0,26)	(+ - 0,55)	(+ - 0,43)	(+ - 0,01)	(+ - 0,02)	(+ - 0,09)	(+ - 0,02)
4	11,07	87,59	89,24	88,72	89,24	89,30	89,09	88,75	88,90	89,16	89,32
		(+ - 0,74)	(+ - 0,55)	(+ - 0,67)	(+ - 0,30)	(+ - 0,03)	(+ - 0,23)	(+ - 0,17)	(+ - 0,02)	(+ - 0,04)	(+ - 0,05)
5	13,08	88,21	89,05	88,18	88,81	88,97	89,24	89,31	89,19*	89,19*	88,85
		(+ - 0,47)	(+ - 0,32)	(+ - 1,24)	(+ - 0,29)	(+ - 0,02)	(+ - 0,39)	(+ - 0,28)	(+ - 0,11)	(+ - 0,07)	(+ - 0,20)
6	13,83	87,56	88,26	88,37	88,10	88,50	88,26	88,29	88,36	88,03	88,01
		(+ - 0,24)	(+ - 0,26)	(+ - 0,41)	(+ - 0,11)	(+ - 0,27)	(+ - 0,08)	(+ - 0,40)	(+ - 0,24)	(+ - 0,04)	(+ - 0,32)
7	14,15	86,07	86,63	86,38	87,20*	86,72	87,14*	87,29*	87,32*	87,29*	86,65
		(+ - 0,72)	(+ - 0,03)	(+ - 0,27)	(+ - 0,20)	(+ - 0,36)	(+ - 0,28)	(+ - 0,16)	(+ - 0,10)	(+ - 0,12)	(+ - 0,12)
8	14,13	84,53	86,36	85,47	86,02	85,88	86,54*	86,71*	86,55*	85,91	86,28
		(+ - 0,45)	(+ - 0,32)	(+ - 0,23)	(+ - 0,21)	(+ - 0,55)	(+ - 0,09)	(+ - 0,04)	(+ - 0,01)	(+ - 0,27)	(+ - 0,08)
Média	11,42	87,72	88,65	88,18	88,39	88,58	88,65	88,82	88,92	88,63	88,56

Rendimento fermentativo (%) obtido utilizando a média de duas réplicas biológicas. O valor do desvio padrão é apresentado entre parênteses. * Rendimento significativamente superior em relação ao isolado referência (JAY270).



Figura 1. Passos utilizados para a construção do plasmídeo OVP4.

A: Plasmídeo pFA6KanMX4. B: Região promotora do gene T1 com os iniciadores contendo os sítios de corte das enzimas *EcoRI* e *SacII*. C: Mapa físico do plasmídeo integrativo OVP4, que possui a região promotora do gene T1 (pT1). OVOX e OVOY representam iniciadores com sua determinada orientação para amplificação da sequência contendo o gene repórter (*KanMX4*), a região promotora e a região de homologia para recombinação.

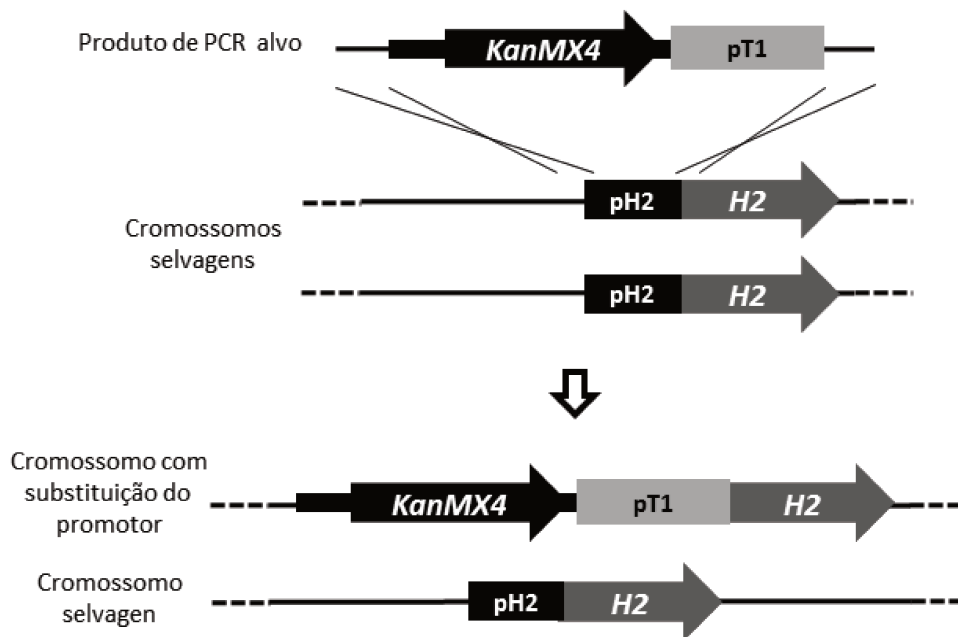


Figura 2. Alteração da região promotora de um gene de interesse.

Após amplificação via PCR da região promotora ligada ao gene de resistência ao antibiótico, o produto foi utilizado na transformação das leveduras utilizando uma região de 42 pares de base de homologia. Células transformadas integraram o produto de PCR via recombinação homóloga em uma cópia do cromossomo.

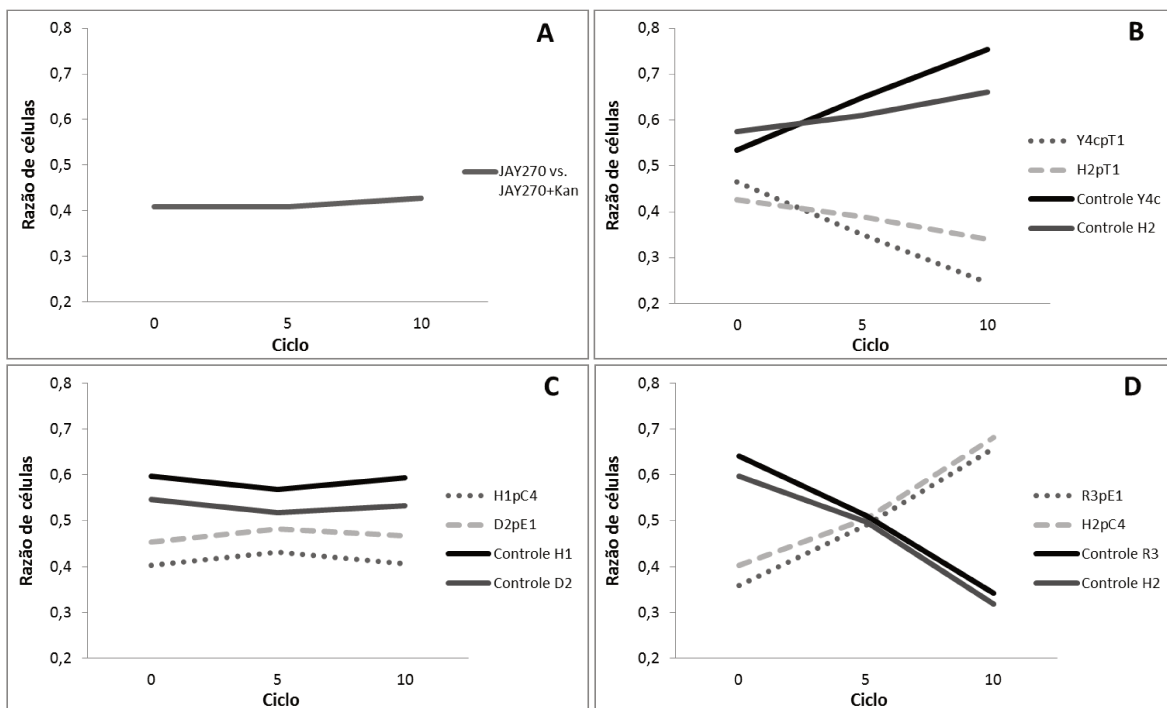


Figura 3. Razão do número de células identificadas nos ensaios de fermentação sob competição.

Nos gráficos são apresentadas as razões do número de células observadas em três pontos durante o ensaio de competição para a amostra controle (A), para linhagens com a região promotora alterada com perfil de crescimento negativo (B), neutro (C) e positivo (D). Para o cálculo dos valores das razões de células de cada linhagem, o número de colônias presentes na placa com genética foi dividido pelo número de colônias na placa sem antibiótico. Em cada competição foi utilizado o isolado selvagem JAY270/PE-2 sem marca de seleção (linhas contínuas). As linhagens modificadas são representadas com linhas pontilhadas.

CONCLUSÕES GERAIS

Visando o incremento das características fermentativas de PE-2, nosso grupo propôs um ciclo de estudos para compreensão e manipulação da linhagem. Este ciclo foi dividido em quatro fases: i- sequenciamento do genoma (Argueso et al., 2009, Capítulo 1); ii- desenvolvimento de uma plataforma biológica domesticada auxotrófica para uracila (Galzerani et al., 2010, Capítulo 3); iii- análise do seu transcriptoma em condições industriais (Capítulo 4); e, finalmente, iv- utilização destas informações para o desenvolvimento de uma linhagem superior derivada de PE-2 (Capítulo 5).

Os dados gerados neste trabalho possibilitaram compreender a estrutura e a maquinaria genética de uma das principais linhagens utilizadas no Brasil na produção de bioetanol (PE-2). A obtenção dos dados utilizando modernos equipamentos e técnicas de biologia molecular possibilitaram que um elevado conteúdo de informações fosse detalhadamente reportado. Estas informações foram utilizadas com sucesso em ensaios de manipulação da linhagem, obtendo uma linhagem domesticada e outras com desempenhos fermentativos superiores em comparação ao isolado selvagem. Dessa forma, os resultados apresentados neste trabalho permitiram concluir que:

- A plasticidade do genoma de PE-2, concentrada em suas regiões sub-teloméricas, é uma estratégia utilizada pela levedura para se adaptar mais prontamente aos diferentes estresses impostos durante o processo fermentativo, fornecendo assim uma vantagem competitiva à linhagem.
- Foi identificado também um elevado grau de polimorfismos na linhagem PE-2, tanto em referência a linhagem padrão de laboratório, S288c, quanto entre seus cromossomos homólogos.
- A acumulação de massa celular de PE-2 e a produção de etanol são características independentes e herdadas de maneira semi-dominante e dominante, respectivamente.
- O conjunto de marcadores moleculares desenvolvido a partir de informações do genoma de PE-2 mostrou-se mais eficiente, rápido e menos laborioso que o principal método (PFGE) utilizado atualmente para identificação de linhagens de *Saccharomyces cerevisiae*.
- Os marcadores moleculares foram altamente informativos, inclusive para linhagens utilizadas em processos tecnológicos semelhantes e com mesma origem geográfica.

- A estratégia de domesticação de PE-2 gerou um isolado auxotrófico para uracila sem que as características de robustez associadas à linhagem fossem perdidas.
- A ação de *Lactobacillus fermentum* foi responsável pelo fenômeno de floculação (co-agregação) observado em PE-2 sob condições industriais. Além disso, os ácidos orgânicos produzidos pela bactéria reprimiram significativamente o metabolismo da levedura.
- As diferenças na expressão de genes relacionados com a síntese de vitaminas, principalmente B1, ilustraram a importância da variação do número de cópias desses genes no genoma de PE-2. Através destas características, é possível inferir que estes genes devam ter papel fundamental no processo de adaptação nas condições industriais.
- O alto grau de polimorfismos no genoma de PE-2 permitiu que diferenças na expressão entre os alelos fossem avaliadas nas fermentações industriais. Concluiu-se assim que os padrões de expressão são tanto dependentes como também independentes das condições analisadas.
- Os dados de expressão alélica diferencial permitiram concluir também que a linhagem comercial de PE-2 analisada perdeu o padrão de heterozigosidade no braço direito do cromossomo 13.
- O aumento da expressão de determinados genes ao longo da fermentação, através de manipulações genéticas pontuais, contribuíram para o incremento do desempenho fermentativo em linhagens derivadas de PE-2, principalmente em condições com teores alcoólicos elevados.

REFERÊNCIAS BIBLIOGRÁFICAS

- 't Hoen PAC, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, Boer JM, van Ommen G-JB, den Dunnen JT. 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Research* **36**:e141.
- Abbott DA, Zelle RM, Pronk JT, van Maris AJA. 2009b. Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. *FEMS Yeast Research* **9**:1123-36.
- Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G. 2006. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science (New York, N.Y.)* **314**:1565-8.
- Amorim HV, Lopes ML, de Castro Oliveira JV, Buckeridge MS, Goldman GH. 2011. Scientific challenges of bioethanol production in Brazil. *Applied Microbiology and Biotechnology* **91**:1267-75.
- Argueso JL, Pereira GAG. 2010. Perspective: Indigenous sugarcane yeast strains as ideal biological platforms for the delivery of next generation biorefining technologies. *International Sugar Journal* **112**:86-89.
- Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OVC, Missawa SK, Galzerani F, Costa GGL, Vidal RO, Noronha MF, Dominska M, Andrietta MGS, Andrietta SR, Cunha AF, Gomes LH, Tavares FCA, Alcarde AR, Dietrich FS, McCusker JH, Petes TD, Pereira GAG. 2009. Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Research* **19**:2258-70.
- Babrzadeh F, Jalili R, Wang C, Shokralla S, Pierce S, Robinson-Mosher A, Nyren P, Shafer RW, Basso LC, de Amorim HV, de Oliveira AJ, Davis RW, Ronaghi M, Gharizadeh B, Stambuk BU. 2012. Whole-genome sequencing of the efficient industrial fuel-ethanol fermentative *Saccharomyces cerevisiae* strain CAT-1. *Molecular Genetics and Genomics* **287**:485-94.
- Backhus LE, DeRisi J, Bisson LF. 2001. Functional genomic analysis of a commercial wine strain of *Saccharomyces cerevisiae* under differing nitrogen conditions. *FEMS Yeast Research* **1**:111-125.
- Basso LC, de Amorim HV, de Oliveira AJ, Lopes ML. 2008. Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Research* **8**:1155-63.
- Basso LC, Basso TO, Rocha SN. 2010. Ethanol production in Brazil: The industrial process and its impact on yeast fermentation. In: Bernardes, MAS, editor. *Biofuel Production – Recent Developments and Prospects* 1st ed. Rijeka: InTech, Vol. 1530, p. 596.
- Berretta J, Morillon A. 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Reports* **10**:973-982.

- Blieck L, Toye G, Dumortier F, Verstrepen KJ, Delvaux FR, Thevelein JM, van Dijck P. 2007. Isolation and characterization of brewer's yeast variants with improved fermentation performance under high-gravity conditions. *Applied and Environmental Microbiology* **73**:815-24.
- Bracmort K. 2012. Meeting the Renewable Fuel Standard (RFS) Mandate for Cellulosic Biofuels : Questions and Answers.
- Brown TA. 2002. Genomes 2nd ed. Oxford, UK: BIOS Scientific Publishers, Ltd.
- Cordier H, Mendes F, Vasconcelos I, François JM. 2007. A metabolic and genomic study of engineered *Saccharomyces cerevisiae* strains for high glycerol production. *Metabolic Engineering* **9**:364-378.
- Costa V, Angelini C, de Feis I, Ciccodicola A. 2010. Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine Biotechnology* **2010**:853916.
- DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**:680-686.
- Dequin S. 2001. The potential of genetic engineering for improving brewing, wine-making and baking yeasts. *Applied Microbiology and Biotechnology* **56**:577-588.
- Ehrenreich A. 2006. DNA microarray technology for the microbiologist: an overview. *Applied Microbiology and Biotechnology* **73**:255-73.
- Erasmus DJ, Van Der Merwe GK, van Vuuren HJJ. 2003. Genome-wide expression analyses: Metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress. *FEMS Yeast Research* **3**:375-399.
- Fadiel A, Naftolin F. 2003. Microarray applications and challenges : a vast array of possibilities. *Gene Expression*:1111-1121.
- Galzerani F, Carvalho-Netto OV, Argueso JL, Duarte FM, Missawa S, Pereira GAG. 2010. Processo para produção de uma levedura mutante nulo do gene *URA3* da cepa industrial Pedra-2 (*Saccharomyces cerevisiae*), levedura mutante e uso da mesma. Protocolo n.018100039982.
- Giga-Hama Y, Tohda H, Takegawa K, Kumagai H. 2007. *Schizosaccharomyces pombe* minimum genome factory. *Biotechnology and Applied Biochemistry* **46**:147-155.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 genes. *Science* **274**:546-567.
- Goldemberg J. 2008. The Brazilian biofuels industry. *Biotechnology for Biofuels* **1**:6.

- Halasz G, Van Batenburg MF, Perusse J, Hua S, Lu X-J, White KP, Bussemaker HJ. 2006. Detecting transcriptionally active regions using genomic tiling arrays. *Genome Biology* **7**:59.
- Hirasawa T, Yoshikawa K, Nakakura Y, Nagahisa K, Furusawa C, Katakura Y, Shimizu H, Shioya S. 2007. Identification of target genes conferring ethanol stress tolerance to *Saccharomyces cerevisiae* based on DNA microarray data analysis. *Journal of Biotechnology* **131**:34-44.
- James TC, Campbell S, Donnelly D, Bond U. 2003. Transcription profile of brewery yeast under fermentation conditions. *Journal of Applied Microbiology* **94**:432-48.
- Juneau K, Palm C, Miranda M, Davis RW. 2007. High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proceedings of the National Academy of Sciences of the United States of America* **104**:1522-1527.
- Leite RCC. 2009. Bioetanol combustível: uma oportunidade para o Brasil. Ed. Luís Augusto Barbosa Cortez, Manoel Regis Lima Verde Leal, Marcelo Khaled Poppe, Marcelo Pereira da Cunha 1st ed. Brasilia.
- Li B-Z, Cheng J-S, Qiao B, Yuan Y-J. 2010. Genome-wide transcriptional analysis of *Saccharomyces cerevisiae* during industrial bioethanol fermentation. *Journal of Industrial Microbiology & Biotechnology* **37**:43-55.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* **18**:1509-17.
- Marks VD, Ho Sui SJ, Erasmus D, van der Merwe GK, Brumm J, Wasserman WW, Bryan J, van Vuuren HJJ. 2008. Dynamics of the yeast transcriptome during wine fermentation reveals a novel fermentation stress response. *FEMS Yeast Research* **8**:35-52.
- Meleiro LAC, Maciel-Filho R. 2000. State and parameter estimation based on a nonlinear filter applied to an industrial process control of ethanol production. *Brazilian Journal Chemistry* **17**:991-1002.
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**:1344-9.
- Olesen K, Felding T, Gjermansen C, Hansen J. 2002. The dynamics of the *Saccharomyces carlsbergensis* brewing yeast transcriptome during a production-scale lager beer fermentation. *FEMS Yeast Research* **2**:563-73.
- Oliva-Neto P. 1990. Influência da contaminação por bactérias lácticas na fermentação alcoólica pelo processo de batelada alimentada; Unicamp.
- Oshlack A, Robinson MD, Young MD. 2010. From RNA-seq reads to differential expression results. *Genome Biology* **11**:220.

- Porro D, Gasser B, Fossati T. 2011. Production of recombinant proteins and metabolites in yeasts. *Metabolic Engineering*: 939-948.
- Rossignol T, Dulau L, Julien A, Blondin B. 2003. Genome-wide monitoring of wine yeast gene expression during alcoholic fermentation. *Yeast* **20**:1369-85.
- Shrank S, Farahmand F. 2011. Biofuels regain momentum. *Vital Signs Online*. Washington: Worldwatch Institute.
- da Silva-Filho EA, Brito dos Santos SK, Resende ADM, de Moraes JOF, de Moraes MA, Ardaillon Simões D. 2005. Yeast population dynamics of industrial fuel-ethanol fermentation process assessed by PCR-fingerprinting. *Antonie van Leeuwenhoek* **88**:13-23.
- Stambuk BU, Dunn B, Alves SL, Duval EH, Sherlock G. 2009. Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis. *Genome Research*: 2271-2278.
- Tanaka F, Ando A, Nakamura T, Takagi H, Shima J. 2006. Functional genomic analysis of commercial baker's yeast during initial stages of model dough-fermentation. *Food Microbiology* **23**:717-728.
- Varela C, Cárdenas J, Melo F, Agosin E. 2005. Quantitative analysis of wine yeast gene expression profiles under winemaking conditions. *Yeast* **22**:369-383.
- Wheals AE, Basso LC, Alves DM, Amorim HV. 1999. Fuel ethanol after 25 years. *Trends in Biotechnology* **17**:482-7.
- Wu H, Zheng X, Araki Y, Sahara H, Takagi H, Shimoi H. 2006. Global gene expression analysis of yeast cells during sake brewing. *Applied and Environmental Microbiology* **72**:7353-8.
- Zuzuarregui A, del Olmo M-Í. 2004. Expression of stress response genes in wine strains with different fermentative behavior. *FEMS Yeast Research* **4**:699-710.