

**EVOLUÇÃO DE GENES QUE CONFEREM RESISTÊNCIA A
MÚLTIPLAS DROGAS EM FUNGOS**

CRISTIANI MIRANDA DAVID GOSSANI

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY
RIBEIRO**

CAMPOS DOS GOYTACAZES - RJ.

JULHO, 2014.

**EVOLUÇÃO DE GENES QUE CONFEREM RESISTÊNCIA A
MÚLTIPLAS DROGAS EM FUNGOS**

CRISTIANI MIRANDA DAVID GOSSANI

Dissertação apresentada à
Universidade Estadual do Norte
Fluminense Darcy Ribeiro para
obtenção do título de Mestra em
Biotecnologia e Biotecnologia.

ORIENTADOR: Dr. THIAGO MOTTA VENANCIO

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY
RIBEIRO**

CAMPOS DOS GOYTACAZES - RJ.

JULHO, 2014.

**EVOLUÇÃO DE GENES QUE CONFEREM RESISTÊNCIA A
MÚLTIPLAS DROGAS EM FUNGOS**

CRISTIANI MIRANDA DAVID GOSSANI

Dissertação apresentada à
Universidade Estadual do Norte
Fluminense Darcy Ribeiro para
obtenção do título de Mestra em
Biotecnologia e Biociências.

ORIENTADOR: Dr. THIAGO MOTTA VENANCIO

Aprovada em 07 de Julho de 2014

COMISSÃO EXAMINADORA

Prof. Dr. Alessandro Coutinho Ramos (UENF)

Prof. Dr. Arnaldo Rocha Façanha (UENF)

Prof. Dr. Rodrigo Nunes da Fonseca (UFRJ)

Prof. Dr. Thiago Motta Venancio (UENF)

Cristiani Miranda David Gossani

Agradeço...

Primeiramente a Deus por me agraciar com o dom da vida e por sempre estar ao meu lado, me protegendo e guiando. Obrigada Senhor, por me proporcionar momentos de felicidade, verdade e dor, com os quais pude perceber a grandeza da sua misericórdia e graça, tornando-me assim, um ser humano melhor.

A minha família, meu porto seguro, junto dos quais todas as dificuldades se tornaram leves. Muito obrigada pelo apoio e confiança em todos os momentos e em todas as escolhas.

Ao professor Dr.Thiago Motta Venancio por ter permitido e confiado a realização deste trabalho em seu laboratório. Aos professores, funcionários e colegas do laboratório de Química e Função de Proteínas e Peptídeos (LQFPP), por toda ajuda e apoio.

Ao MS Daniel Bellieny Rabelo pela colaboração, conselhos, paciência e scripts, sem os quais, o desenvolvimento e a conclusão deste trabalho teriam sido muito mais árduos.

À Universidade Estadual do Norte Fluminense Darcy Ribeiro por permitir a realização deste trabalho e a CAPES e FAPERJ pelo apoio financeiro.

Sem a presença de cada uma das pessoas e instituições citadas, com certeza, esse trabalho não teria sido concluído. A todos o meu...

Muito Obrigada!

RESUMO

Apesar de estudos moleculares importantes terem sido realizados com os genes que conferem resistência a múltiplas drogas (*multidrug-resistance genes* - MDRs) a trajetória evolutiva destes genes ainda não é totalmente conhecida. Existem questões importantes que precisam de resposta, como por exemplo: quando e como os MDRs surgiram nos fungos? Como o evento de duplicação de genoma completo impacta a evolução de resistência a novas drogas? Este trabalho apresenta evidências de que: (i) Os MDRs possuem frequência de mutação e mais interações genéticas e proteína-proteína (GIs e PPIs, respectivamente) que os não MDRs; (ii) Os MDRs duplicados (dMDRs) possuem o mesmo número de GIs e PPIs que os MDRs únicos (sMDRs). A frequência de mutação é maior nos dMDRs do que nos sMDRs, indicando maior robustez contra mutações no primeiro grupo, provavelmente devido a redundância genética. A maioria dos dMDRs (59,18%) precedem o surgimento dos fungos enquanto que somente 35,58% dos sMDRs surgiram antes do mesmo filo. Uma porção significativa dos sMDRs (32,88%) é ancestral ao surgimento do clado das *Saccharomyces* enquanto que somente 10,34% dos dMDRs surgiram nesse mesmo ancestral; (iii) Os parálogos de duplicação completa do genoma (WGD) dos MDRs possuem muito menos GIs e PPIs que seus respectivos MDRs-WGD. A mediana da taxa de mutação por gene também é maior nos parálogos WGD do que nos MDRs-WGD. Estes resultados sugerem que os parálogos WGD estão perdendo função ou sofrendo rápida divergência funcional; (iv) Entretanto, os parálogos de pequena escala de duplicação (SSD) dos MDRs possuem um pouco mais de GIs e menos PPIs que seus respectivos MDRs. A mediana da frequência de mutação é maior nos MDRs-SSD do que nos seus parálogos SSD, sugerindo que genes deste último grupo são mais propensos a evoluir novas funções relacionadas à resistência a drogas.

ABSTRACT

In spite of the important molecular studies on the genes that confer multiple drugs resistance (multidrug-resistance genes – MDRs) the evolutionary trajectories of these genes are still unknown. There are important questions that remain unanswered, for example: how and when did MDRs emerge in fungi? Does gene duplication impact the evolution of resistance of new drugs? In this work we show that: (i) MDRs have higher mutation frequency and more, genetic and protein-protein interactions (GIs and PPIs, respectively) than non-MDRs. (ii) Duplicate MDRs (dMDRs) have same number of GIs and PPIs than singleton MDRs (sMDRs). The mutation frequency is greater on dMDRs than on sMDRs, suggesting that the first group has more mutational robustness, probably due to genetic redundancy. The majority of dMDRs (59.18%) predate the emergence of fungi, while only 35.58% of sMDRs emerged before same phylum. A significant portion of sMDRs (32.88%) is ancestral to emergence of *Saccharomyces* clade while solely 10.34% of dMDRs emerged in this same ancestral. (iii) The paralogs from whole genome duplication (WGD) of MDRs have much less GIs and PPIs than its respective MDRs-WGD. The mutation ratio median by gene is also greater on WGDs paralogs than its MDRs-WGD. These findings suggest that WGD paralogs are losing functions or undergoing rapid functionally divergence: (iv) Nevertheless the small scale duplication (SSD) paralogs from the MDRs have slightly more GIs and less PPIs than its respective MDRs-SSD. Mutation frequency median is greater in MDRs-SSD than in its SSD paralogs, suggesting that the last group is more likely to evolve novel functions related to drug resistance.

ABREVIações

CPnet – Rede de fenótipos químicos (*chemical-phenotype network*)

dMDRs – MDRs (vide abaixo) duplicados

GIs – Interações genéticas (*genetic interactions*)

MDRs – Genes que conferem resistência a múltiplas-drogas (*multidrug-resistance genes*)

MDRs-SSD – MDRs de pequena escala de duplicação do genoma

MDRs-WGD – MDRs de duplicação completa do genoma

PARs-SSD – Parálogos SSD de MDRs-SSD

PARs-WGD - Parálogos WGD de MDRs-WGD

PPIs – Interações proteína-proteínas (*protein-protein interactions*)

sMDRs – MDRs únicos (*single MDRs*)

SSD – Pequena escala de duplicação (*small-scale duplication*)

WGD – Duplicação completa do genoma (*whole-genome duplication*)

GLOSSÁRIO

Arquitetura de domínio diferente – Ocorre quando proteínas duplicadas (parálogas) possuem domínios protéicos em posições ou quantidades diferentes em uma das cópias.

BLASTClust – Programa autônomo dentro do pacote BLAST usado para agrupar sequências de nucleotídeos ou proteínas para obtenção de grupo de sequências não redundantes.

BLASTP – *Basic Local Alignment Search Tool Protein* é um algoritmo para comparação de sequências de proteínas, que permite comparar uma sequência específica com um banco de dados de sequências de proteínas e identificar sequências da biblioteca que se assemelham a sequência em questão a partir de um limiar.

Bit score – Valor normalizado expresso em *bits* que permite estimar a magnitude do espaço de pesquisa. Ele informa a qualidade do alinhamento, quanto mais alto o valor melhor é o alinhamento.

E-value – O valor esperado (E-value) representa o número de diferentes alinhamentos com pontuação equivalente ou melhor que um alinhamento esperado ocorrer em uma pesquisa de banco de dados ao acaso. Quanto menor o E-value mais significativa é a pontuação e o alinhamento.

HMMER – O HMMER é um software utilizado para pesquisar sequências em bancos de dados de sequências de proteínas homólogas e para fazer alinhamento de sequências. O método utiliza modelos probabilísticos chamados *profile hidden Markov models*.

Hubs – Termo utilizado para os genes mais conectados de uma rede.

MCL – O **Markov Cluster Algorithm** é um algoritmo que permite agrupar sequências similares de proteínas em famílias protéicas.

Mediana da fixação da mutação - Mediana da frequência de mutação de um conjunto de genes na última geração da população em estudo.

Mediana da frequência de mutação - Mediana da frequência de mutação de um conjunto de genes em cada geração da população em estudo.

Mediana da taxa de mutação – Mediana da taxa de mutação de um conjunto de genes.

Pfam – O banco de dados Pfam é uma grande coleção de famílias protéicas, cada uma representada por múltiplos alinhamentos de sequências e *hidden Markov models*.

Índice

1- Introdução.....	12
1.1- Fungos patogênicos e <i>Saccharomyces cerevisiae</i>	12
1.2- Quimiogenética e quimiogenômica.....	14
1.3- Análise integrativa dos dados de quimiogenética e MDRs.....	15
1.4- Evolução Genômica.....	16
1.5- Divergência Funcional.....	18
2- Justificativas.....	20
3- Objetivos.....	20
3.1- Objetivo Geral.....	20
3.2- Objetivos Específicos.....	20
4- Metodologia.....	21
4.1- Duplicação dos parálogos WGD e SSD.....	21
4.2- Dados de Quimiogenética.....	21
4.3- Interações Genéticas e Proteína-Proteínas.....	21
4.4- Mutações Através das Gerações.....	22
4.5- Análise de Sequências e Agrupamentos de proteínas.....	22
4.6- Análises Estatísticas.....	24
5- Resultados.....	24
5.1- Agrupamento dos MDRs de <i>S. cerevisiae</i> e análise dos domínios protéicos.....	24
5.1.1- Muitos MDRs de <i>S. cerevisiae</i> surgiram por eventos de duplicação.....	24
5.1.2- A maioria dos MDRs de <i>S. cerevisiae</i> possuem domínios protéicos conservados idênticos aos seus parálogos.....	25
5.2- Análise do surgimento de mutações em MDRs através das gerações em <i>S. cerevisiae</i>	27
5.2.1- Os MDRs são proporcionalmente mais representados entre os “putative drivers” do que os não MDRs.....	27
5.2.2- Os MDRs sofrem mais mutações que os não MDRs.....	28
5.2.3- A frequência de mutação dos dMDRs é maior do que os sMDRs.....	29
5.2.4- Os MDRs sofrem menos mutações que os seus parálogos.....	30
5.2.5- As mutações através das gerações são similares entre MDRs-WGD e seus parálogos WGD.....	31
5.2.6- OS MDRs-SSD com parálogos SSD possuem maior frequência de mutação.....	32
5.2.7- Os MDRs-WGD e os MDRs-SSD possuem frequência de mutação similar.....	33
5.3- Análise do perfil funcional dos MDRs de <i>S. cerevisiae</i>	34
5.3.1- Os MDRs possuem maior diversidade funcional do que os não MDRs.....	34
5.3.3- Os dMDRs possuem maior diversidade funcional que os seus parálogos.....	35
5.3.4- Os parálogos WGD dos MRDs-WGD estão perdendo função após a duplicação.....	36
5.3.5- Os MDRs-SSD e seus parálogos SSD estão ganhando novas funções após a duplicação.....	37
5.3.6- Os MDRs-WGD e os MDRs-SSD possuem números similares de interações protéica e genética.....	38
5.4- Análise da classificação taxonômica dos MDRs.....	39
5.4.1- Os MDRs são mais ancestrais do que os não MDRs.....	39
5.4.2- Os dMDRs são muito mais conservados do que os sMDRs.....	40

5.4.3- Os dMDRs são mais conservados que seus parálogos.....	41
5.4.4- Os MDRs-WGD e seus parálogos WGD possuem conservação similar através dos taxa.....	42
5.4.5- Os MDRs-SSD são mais conservados que os seus parálogos.....	43
5.4.6- As famílias protéicas com MDRs-SSD são muito mais conservadas do que as famílias protéicas com MDRs-WGD.....	44
6- Discussão.....	46
7- Conclusão.....	49
8- Referências Bibliográficas.....	51
9- Apêndice.....	59

1- Introdução

1.1- Fungos patogênicos e *Saccharomyces cerevisiae*

Os fungos constituem um grande grupo monofilético de eucariotos, caracterizados por crescimento osmotrófico, no qual enzimas são secretadas para converter polissacarídeos e proteínas em carboidratos mais simples e aminoácidos, que são internalizados e utilizados no crescimento. Existem duas formas morfológicamente distintas de fungos: levedura unicelular (a qual cresce por brotamento ou fissão simples) e fungo filamentosos, que produz hifas que se agregam para formar uma rede chamada micélio. A natureza osmotrófica é extremamente eficiente na colonização de diversos habitats, fazendo dos fungos os principais decompositores de biomassa existentes (de Boer *et al.*, 2005).

Existem aproximadamente 1,5 milhões de espécies de fungos conhecidas, sendo sua vasta maioria de natureza não-patogênica e saprofítica. Algumas espécies no entanto causam doenças em animais, como por exemplo *Batrachochytrium dendrobatidis*, que infecta diversos anfíbios (Pounds *et al.*, 2006). Há ainda espécies de fungos fitopatogênicos produtores de micotoxinas, que devastam plantações e trazem grandes prejuízos para agricultura (ex: *Fusarium* spp) (Egan e Talbot, 2008). Doenças fúngicas humanas se tornaram um problema sério a partir do século XX, devido ao aumento: da eficiência do tratamento de infecções bacterianas; do número de pacientes imunodeficientes; e do uso de dispositivos médicos de longa permanência (Butler, 2010). A maioria dos fungos patogênicos humanos pertencem aos filos Ascomycota e Basidiomycota (Figura 1), dentre eles *Candida albicans*, *Cryptococcus neoformans* e *Aspergillus fumigatus* (Butler, 2010).

Pertencente ao sub-filo Saccharomycotina, *Saccharomyces cerevisiae* é um excelente organismo modelo e devido ao fato de ter sofrido uma duplicação integral do genoma (Kellis *et al.* 2004; Wolfe e Shields, 1997), essa levedura vem sendo um dos principais organismos utilizados em estudos sobre duplicação gênica. Além do fácil cultivo e facilidade de transformação, vale ainda ressaltar a existência de bibliotecas de mutantes individuais para praticamente todos os genes preditos no genoma (Giaever *et al.*, 2002; Winzeler *et al.*, 1999), possibilitando a experimentação em escala genômica de forma eficiente. Atualmente existem quatro coleções de

mutantes de *S. cerevisiae*: uma diploide homocigota para genes não essenciais, uma diploide heterocigota que contém tanto genes essenciais quanto não essenciais e duas haploides derivadas uma de cada coleção diploide (Scherens e Goffeau, 2004). Estas bibliotecas já foram utilizadas para definir o conjunto de genes essenciais para o crescimento em meio de cultura rico (Giaver *et al.*, 2002), bem como sob diversas condições de estresse ambiental e químico (Hillenmeyer *et al.*, 2008; Parsons *et al.*, 2006).

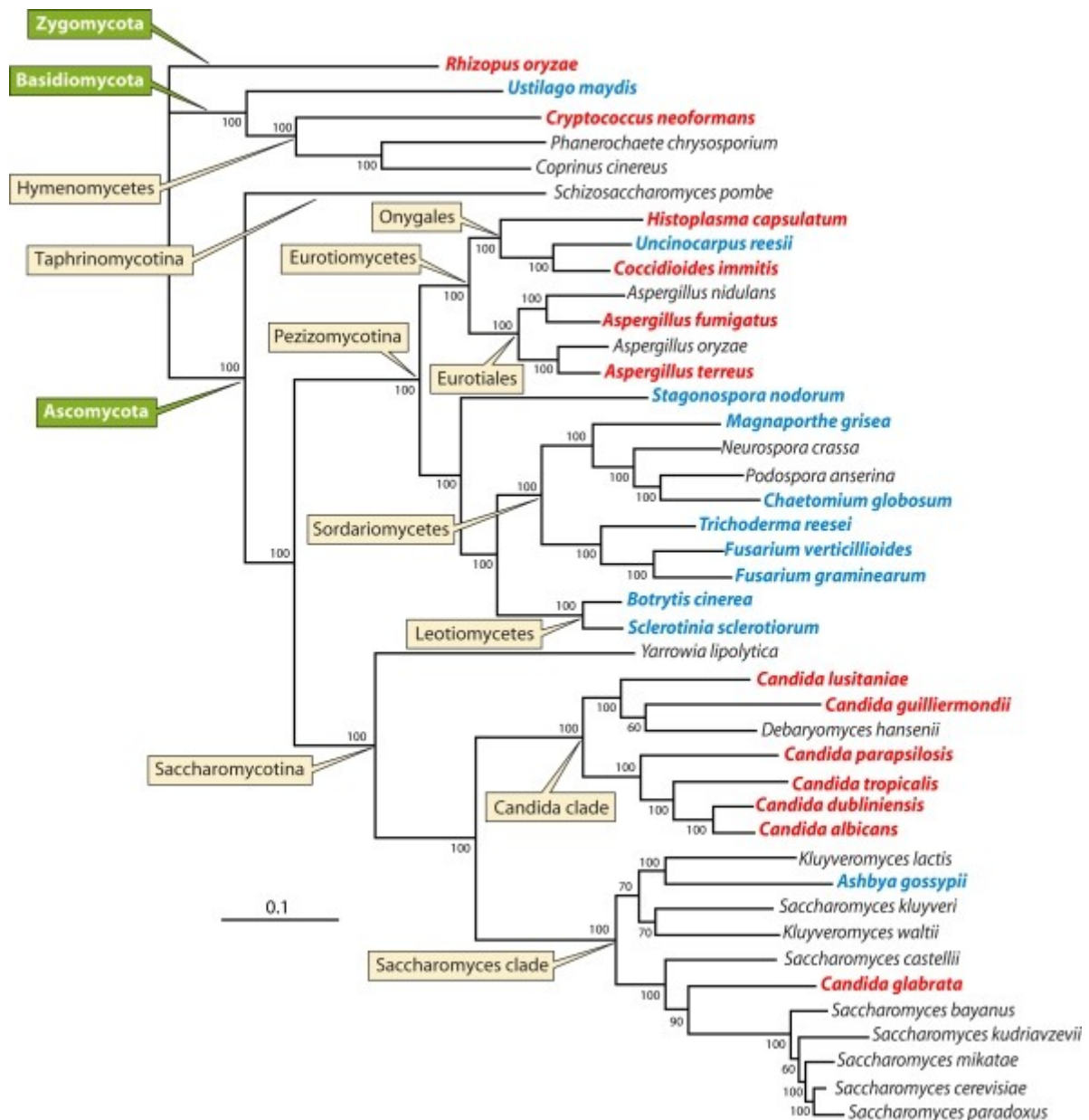


Figura 1: **Filogenia dos fungos.** A filogenia por maximum-likelihood mostra 42 fungos com genomas completamente sequenciados. Fungos fitopatogênicos estão em azul e patógenos humanos estão em vermelho. A porcentagem do Bootstrap para cada nó está representada na base de cada ramo. Fonte: Butler, 2010.

1.2- Quimiogenética e quimiogenômica

Em geral, o perfil de sensibilidade química das cepas de *S. cerevisiae* mutantes pode ser estudado usando abordagens de “*chemical genetics*” ou quimiogenética. Atualmente as três principais metodologias são: (1) linhagens mutantes cultivadas individualmente (ensaio não competitivo) em microplacas em meio de cultura líquido tendo sua taxa de crescimento monitorada por espectrofotometria. As curvas de crescimento na presença e ausência de um dado composto bioativo são comparadas para avaliar a sensibilidade do mutante ao composto (Warringer *et al.*, 2003; Engler *et al.*, 1999) (Figura 2 A); (2) na segunda abordagem cada cepa tem um *barcode* no local do gene deletado, que identifica a cepa e permite o cultivo de toda a biblioteca no mesmo recipiente (ensaio competitivo) (i.e. *in batch*). A biblioteca é então cultivada na presença e ausência de um dado composto químico e as curvas de crescimento estimadas a partir da utilização de um microarray que quantifica cada *barcode* (Ooi *et al.*, 2003). Caso um determinado *barcode* apresente-se mais abundante na ausência de uma determinada droga, infere-se que o gene deletado está associado com a tolerância ou sobrevivência à substância em questão (Figura 2 B); (3) a terceira técnica compreende o crescimento de cepas com genes deletados arranjadas em meio sólido na presença e na ausência do composto de interesse (Parsons *et al.*, 2004; Davis-Kaplan *et al.*, 2004). A taxa de crescimento das colônias é estimada pelo tamanho da colônia comparado relativo ao controle (Figura 2 D). Apesar das vantagens e limitações de cada técnica, o uso de *barcodes* é atualmente o que se tem de melhor, compreendendo o estado da arte em quimiogenética (Lee *et al.*, 2014).

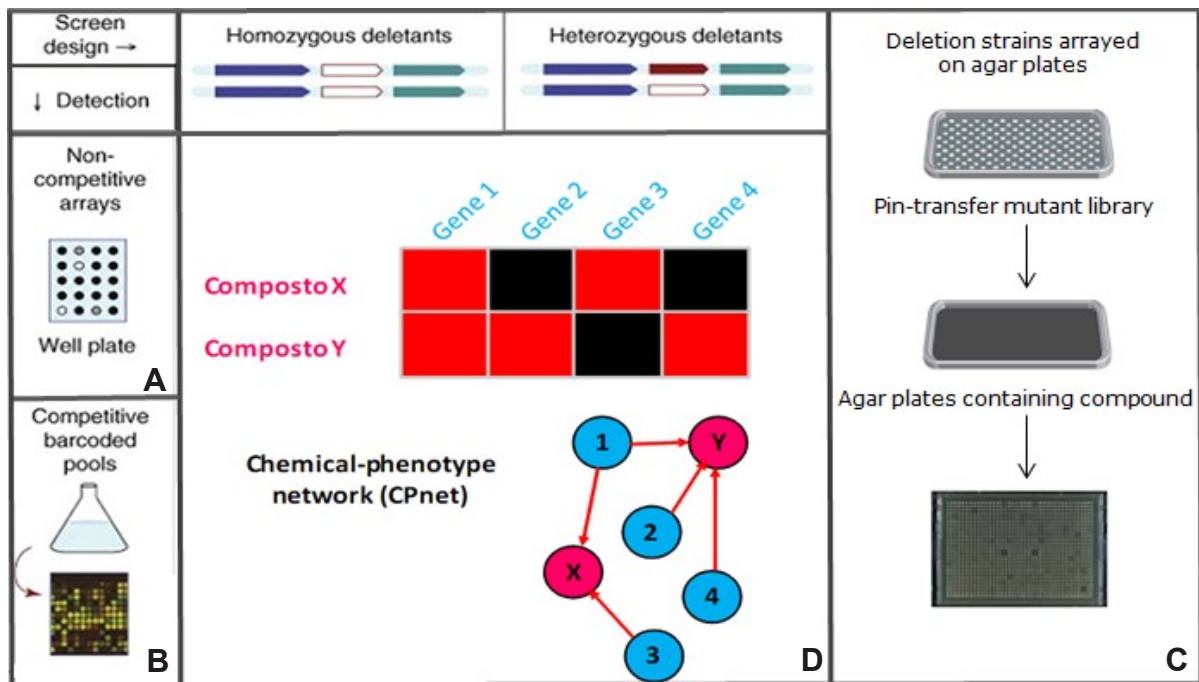


Figura 2: **Representação esquemática de experimentos de chemical genetics.** O crescimento de células mutantes de um dado gene é comparado na presença e ausência de um composto químico em meio líquido, podendo o crescimento ser não-competitivo (A) ou competitivo (B). O crescimento de células mutantes do gene em questão também pode ser comparado com o crescimento das células selvagens ambos na presença do composto químico em meio sólido (C). Os resultados podem ser representados como um heatmap (figura superior) ou rede (figura inferior). Em vermelho (heatmap) e em azul (rede) estão os genes que quando mutados e na presença do composto x ou y em rosa (heatmap, rede) reduziram o crescimento da *S. cerevisiae* e em preto (heatmap) os genes mutados que não interferiram no crescimento da levedura na presença dos compostos. (D). Os painéis acima do quadro D representam: a esquerda três genes duplicados, o par de gene central sofreu deleção homozigótica; e a direita três genes duplicados, o par de gene central uma deleção heterozigótica. Fonte: Adaptado de Kapitzky et al., 2010.

1.3- Análise integrativa dos dados de quimiogenética e MDRs

Utilizando os dados dos experimentos de quimiogenética, Venancio *et al.* (2010) realizou uma análise integrativa da exposição de mutantes a agentes químicos, onde foram coletados dados de 34 publicações, que foram usados para montar uma rede conectando genes e agentes químicos, que foi chamada **chemical-phenotype network** (Cpnet). Análises da topologia desta rede revelaram diversos mecanismos envolvidos na tolerância natural contra uma gama de substâncias de diferentes características físico-químicas (Venancio *et al.*, 2010). Verificou-se inclusive que genes parálogos podem estar envolvidos na tolerância a substâncias totalmente diferentes e que substâncias com um mesmo mecanismo de ação (e.g.

dano ao DNA) nem sempre são detoxificadas pela mesma via metabólica (Venancio *et al.*, 2010).

Uma das principais análises feitas em redes biológicas é a identificação dos genes mais conectados (*i.e.* hubs). Foram identificados 517 *hubs* na Cpnet, sendo estes genes necessários ao crescimento na presença de vários compostos químicos. Estes hubs foram então chamados “*multidrug resistance genes*” (MDRs). Entre os MDRs são encontrados transportadores de membrana, fatores de transcrição e proteínas envolvidas no tráfico endossomal (Venancio *et al.*, 2010; Hillenmeyer *et al.*, 2008). Os MDRs tendem a ser capacitores fenotípicos – genes que quando afetados resultam na perda de robustez da morfologia celular contra modificações genéticas e ambientais (Venancio *et al.*, 2010; Levy e Siegal, 2008). Além disso, MDRs tipicamente não possuem TATA-box em suas regiões promotoras, sendo preferencialmente regulados por TFIID, constituídos de fatores que estão associados com expressão constitutiva e com pouco ruído (Basehoar *et al.*, 2004).

1.4- Evolução Genômica

As relações de homologia entre genes, são tipicamente definidas em: ortólogos, que são genes em espécies diferentes que surgem a partir de um único gene do ancestral comum da espécie por um processo de especiação; e parálogos que são genes na mesma espécie que surgiram a partir de um gene no ancestral da espécie por um processo de duplicação (Figura 3) (Searls, 2008). A identificação dos ortólogos é indispensável para a reconstrução evolutiva dos genomas das espécies (Gabaldón e Koonin, 2013). Parálogos são creditados como sendo a principal fonte de diversidade funcional e conseqüentemente possuem grande importância em processos adaptativos (Fares *et al.*, 2013). Os genes podem ser duplicados de duas maneiras: a partir de duplicação completa do genoma (WGD) e/ou duplicação por pequena escala (SSD). A origem dos genes duplicados (WGD ou SSD) possui papel chave no destino do gene (Lynch *et al.*, 2001). Estudos de genética populacional indicam que a maioria dos genes duplicados retornam ao status inicial de gene único após a duplicação, pois uma cópia acumulará mutações deletérias até a sua completa erosão (Conant e Wolfe, 2008; Lynch *et al.*, 2001). Os genes permanecerão no genoma se: (i) a duplicação do gene favorecer o organismo

com robustez mutacional (Gu *et al.*, 2003); (ii) existir seleção positiva para o aumento da dosagem do gene (Conant e Wolfe, 2008); (iii) genes duplicados compartilharem as funções do gene ancestral (Force *et al.*, 1999); (iv) surgir uma função completamente diferente e útil em uma das cópias (He e Zhang, 2005); (v) ou se as cópias divergirem no seu perfil de expressão (Gu *et al.*, 2002).

Venancio *et al.* (2010) também analisaram o destino de alguns MDRs parálogos, por exemplo, Ypt31 confere tolerância a 39 compostos químicos, enquanto o seu parólogo Ypt32 está ligado a apenas um composto químico. Evidências como estas sugerem fortemente a divergência funcional entre parálogos recentes (Fares *et al.*, 2013).

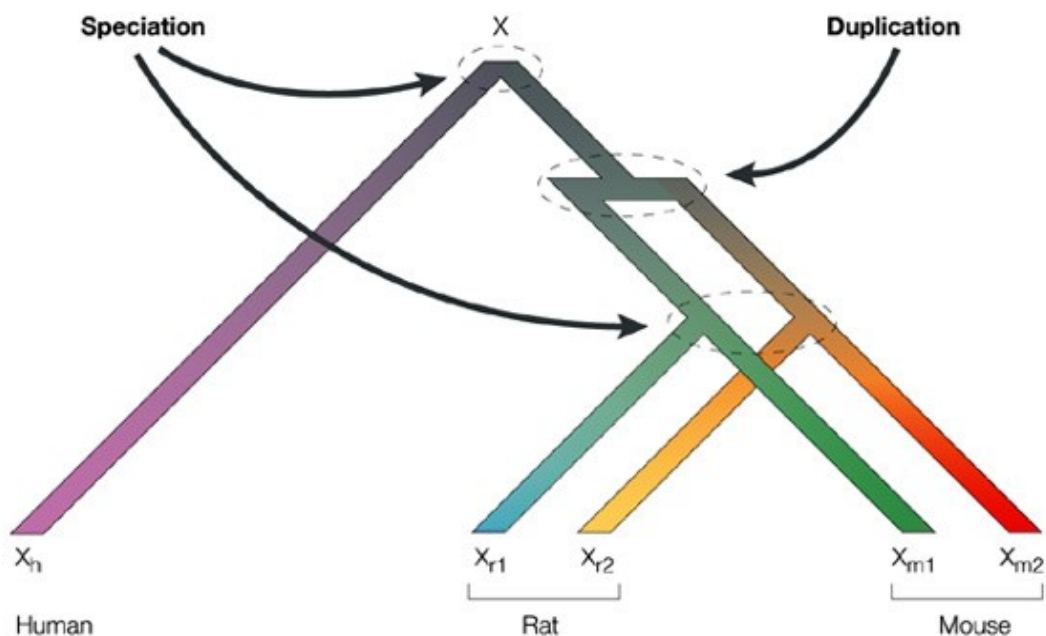


Figura 3: Principais diferenças entre ortólogos e parálogos. Ortólogos são genes em espécies diferentes que surgem a partir de um único gene do ancestral comum da espécie por um processo de especiação. Parálogos são genes na mesma espécie que surgiram a partir de um gene no ancestral da espécie por um processo de duplicação. Na árvore filogenética acima, o gene X sofre uma duplicação seguida de um evento de especiação. Logo, as duas versões dos genes estão presentes nas duas espécies modernas de roedores. Xr1 e Xr2 são parálogos em ratos enquanto Xm1 e Xm2 são parálogos em camundongos. O gene Xh de humano possui dois ortólogos em cada espécie de roedores. Enquanto o Xr1 é ortólogo de Xm1 e Xr2 é ortólogo de Xm2. Fonte: Searls, 2008.

1.5- Divergência Funcional

O perfil das interações quimiogenéticas, genéticas, proteína-proteínas (Figura 4) e a evolução do genoma através das gerações fornecem informações valiosas para a caracterização funcional de redundância e divergência entre parálogos (VanderSluis *et al.*, 2010; Zhang *et al.*, 2005). Zhang *et al.* (2005) analisaram o perfil de interações proteína-proteínas de genes duplicados em levedura e identificaram uma forte correlação positiva entre o nível de conectividade dos genes ancestrais e o número de parceiros de interação compartilhados na rede de interações proteína-proteínas. Recentemente, Fares *et al.* (2013) mostraram que genes duplicados por SSD e WGD possuem diferente destino após a duplicação. Através do mapa das interações genéticas e do ressequenciamento de 27 genomas de *S. cerevisiae* evoluídos por 2.200 gerações eles demonstraram que parálogos SSD sofrem neofuncionalização após duplicação enquanto os parálogos WGD compartilham funções do gene ancestral. O estudo de Lang *et al.* (2013) revela que múltiplas mutações surgem e se movem sincronicamente através da população como mutações que surgem em grupos. Eles identificaram todos os genes mutados através das gerações de levedura por sequenciamento completo dos genomas das populações. Estes resultados são importantes para elucidar se mutações nos MDRs são benéficas e conferem vantagens adaptativas para as populações de levedura. Recentemente, Wisner *et al.* (2013) revelaram que a adaptação pode continuar por longos períodos em organismos assexuais mesmo em ambientes constantes.

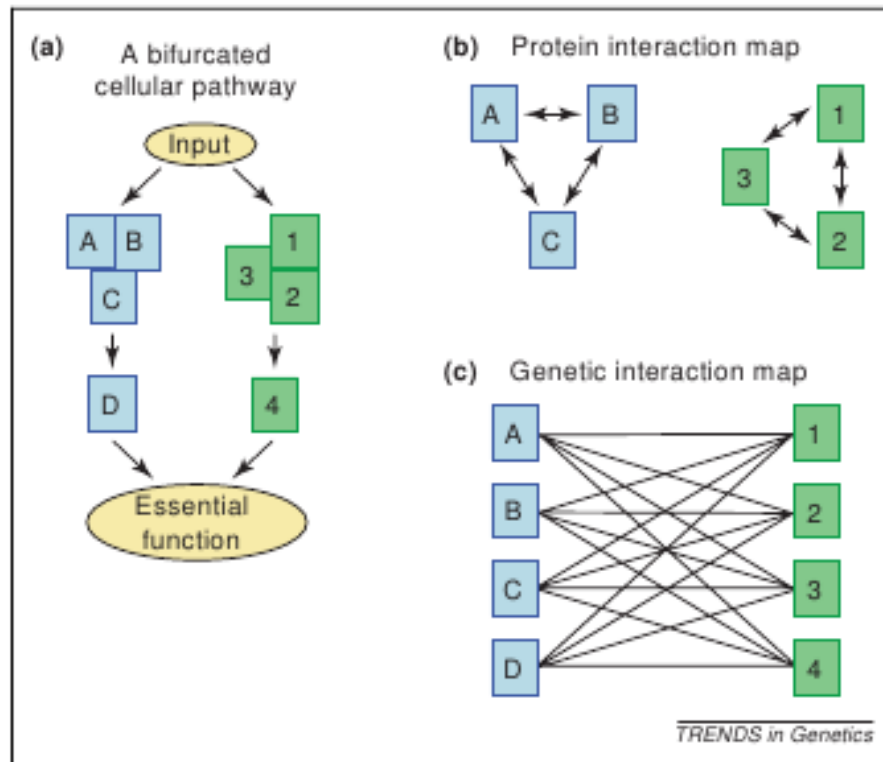


Figura 4: Interações genéticas versus interações proteína-proteínas. (a) Uma via celular bifurcada terá redes distintas nos mapas de interações genéticas e proteína-proteínas. As proteínas A, B, C e D (azul) e as proteínas 1, 2, 3, e 4 (verde) são membros de duas vias funcionalmente redundantes e necessárias para realizar uma função essencial. As proteínas A, B e C interagem fisicamente umas com as outras, assim como as proteínas 1, 2 e 3. (b) O mapa de interações física identifica interações baseada nas interações proteína-proteínas, enquanto o mapa de interações genéticas (c) identifica “interações” baseadas em função sem a necessidade de interação física entre as proteínas. Fonte: Ooi et al., 2005.

2- Justificativas

Os MDRs estão envolvidos em mecanismo de resistência a drogas antifúngicas, que atualmente é um problema agropecuário e de saúde pública (Pfaller, 2012). Apesar de existir estudos sobre as bases genéticas dos MDRs, sua história evolutiva permanece pouco compreendida. Neste projeto, propomos analisar a trajetória evolutiva dos MDRs em várias espécies de fungos, visando essencialmente compreender os fatores que fazem um gene ser recrutado para mecanismos de tolerância a drogas e como esses genes são preservados nas linhagens de fungos.

3- Objetivos

3.1- Objetivo Geral

Analisar a trajetória evolutiva de genes envolvidos em mecanismo de tolerância a drogas em *S. cerevisiae*.

3.2- Objetivos Específicos

1. Selecionar genomas de fungos para análises evolutivas;
2. Analisar ortologia e paralogia nestes genomas agrupando os genes com base na similaridade das proteínas preditas;
3. Integrar estudos mais recentes de quimiogenética aos dados da CPnet;
4. Verificar a conservação dos MDRs de *S. cerevisiae* em outros fungos;
5. Investigar a divergência funcional dos MDRs.

4- Metodologia

4.1- Duplicação dos parálogos WGD e SSD

Os pares de genes duplicados por WGD e SSD foram gentilmente fornecidos por Mario Fares (Fares *et al.*, 2013). Os pares de parálogos foram definidos como o resultado do melhor *hit* recíproco do BLASTP *all-against-all* (E-value < 1E-5 e bit score > 50) (Altschul *et al.*, 1997). Os parálogos foram classificados como WGD resultantes de duplicação completa do genoma que ocorreu na linhagem de *S. cerevisiae* a 100-150 milhões de anos, de acordo com a lista fornecida pelo YGOB (*Yeast Gene Order Browser*, <http://wolfe.gen.tcd.ie/ygob/>) (Byrne e Wolfe, 2005). Todos os pares de genes resultantes do melhor *hit* recíproco do BLASTP *all-against-all* não encontrados na lista do YGOB foram considerados pares parálogos SSD.

4.2- Dados de Quimiogenética

Os dados da CPnet foram obtidos do material suplementar de Venancio *et al.*, 2010 (Additional File 1). A CPnet foi atualizada com nove novas publicações de quimiogenética (Jaime *et al.*, 2012; Peyroche *et al.* 2012; Blackman *et al.*, 2012; Hoon *et al.*, 2011; Castro *et al.*; 2011; Barreto *et al.*, 2011; Uluisik *et al.*, 2011; Alamgir *et al.*, 2010; Dias *et al.*, 2010). Atualmente a CPnet tem 57.701 interações, 5.429 genes e 439 compostos químicos. Apartir desta atualização foram encontrados 63 novos MDRs em *S. cerevisiae*, elevando o número total para 580 MDRs.

4.3- Interações Genéticas e Proteína-Proteínas

Foram utilizadas as últimas atualizações do mapa genético funcional de *S. cerevisiae* (Supplementary file S3 de <http://drygin.cabr.utoronto.ca/~costanzo2009/> - dado acessado em 22/05/2013) (Costanzo *et al.*, 2010). Este mapa funcional é baseado na metodologia de *synthetic genetic array* (Tong *et al.*, 2001), na qual interações genéticas são sistematicamente mapeadas pela produção de mutantes únicos e duplos (Tong *et al.*, 2004). Costanzo *et al.*, (2010) identificaram interações digênicas como aquelas que os mutantes apresentam desvio significativo do crescimento comparado ao efeito do crescimento multiplicativo dos dois mutantes

únicos. Defeitos de crescimento foram estimados de acordo com o tamanho das colônias.

As interações proteína-proteínas foram selecionadas do SGD (*Saccharomyces Genome Database*, <http://www.yeastgenome.org/> - dado acessado em 11/06/2013) (Cherry *et al.*, 2012) o qual foi incorporado do BioGRID (*Biological General Repository for Interaction Datasets*, <http://thebiogrid.org/>) (Stark *et al.*, 2006).

4.4- Mutações Através das Gerações

Dados do sequenciamento do genoma completo de várias populações de *S. cerevisiae* foram recentemente publicado por Lang *et al.*, (2013), que identificaram mutações em 1000 gerações de 40 populações de levedura crescendo em meio rico. Cada população foi sequenciada 100 vezes em 12 pontos ao longo do tempo (aproximadamente a cada 80 gerações) totalizando 480 pontos. Foram selecionados dados de 723 regiões codificadoras de proteína a partir da tabela suplementar <http://www.nature.com/nature/journal/v500/n7464/full/nature12344.html#supplementary-information> – dados acessados em 17/10/2013).

4.5- Análise de Sequências e Agrupamentos de proteínas

Foram selecionados 40 proteomas preditos de fungos e três de protistas como membros externos ao filo de interesse (Apêndice). Para análise dos genomas foi utilizado o BLASTCLUST (www.ncbi.nlm.nih.gov) com limiar de similaridade de 0,89 e tamanho mínimo de cobertura de 0,9. Estes parâmetros foram empregados para remover eventuais isoformas de *splicing* de todas as proteínas de cada genoma com exceção dos genomas de *S. cerevisiae* e *T. vaginalis* para as quais foram utilizadas proteínas do RefSeq. O resultado do BLASTCLUST foi submetido à análise usando BLASTP (Altschul *et al.*, 1997) *all-versus-all* com limiar de E-value < 1E-5. Os resultados do BLASTP alimentaram o *Markov Cluster Algorithm* (MCL) (*inflation* = 3) (van Dongen, 2000). Em seguida as famílias proteicas contendo as proteínas agrupadas por similaridade de sequência foram processadas com *scripts* desenvolvidos na linguagem Perl (Schwartz *et al.*, 2001) (Figura 5).

Os domínios conservados das proteínas foram identificados utilizando HMMER (Eddy, 2009) e Pfam-A 27,0, um banco de dados de domínios conservados

(Punta *et al.*, 2012), para todas as sequências de MDRs de *S. cerevisiae* e seus respectivos parálogos (E-value < 0,01).

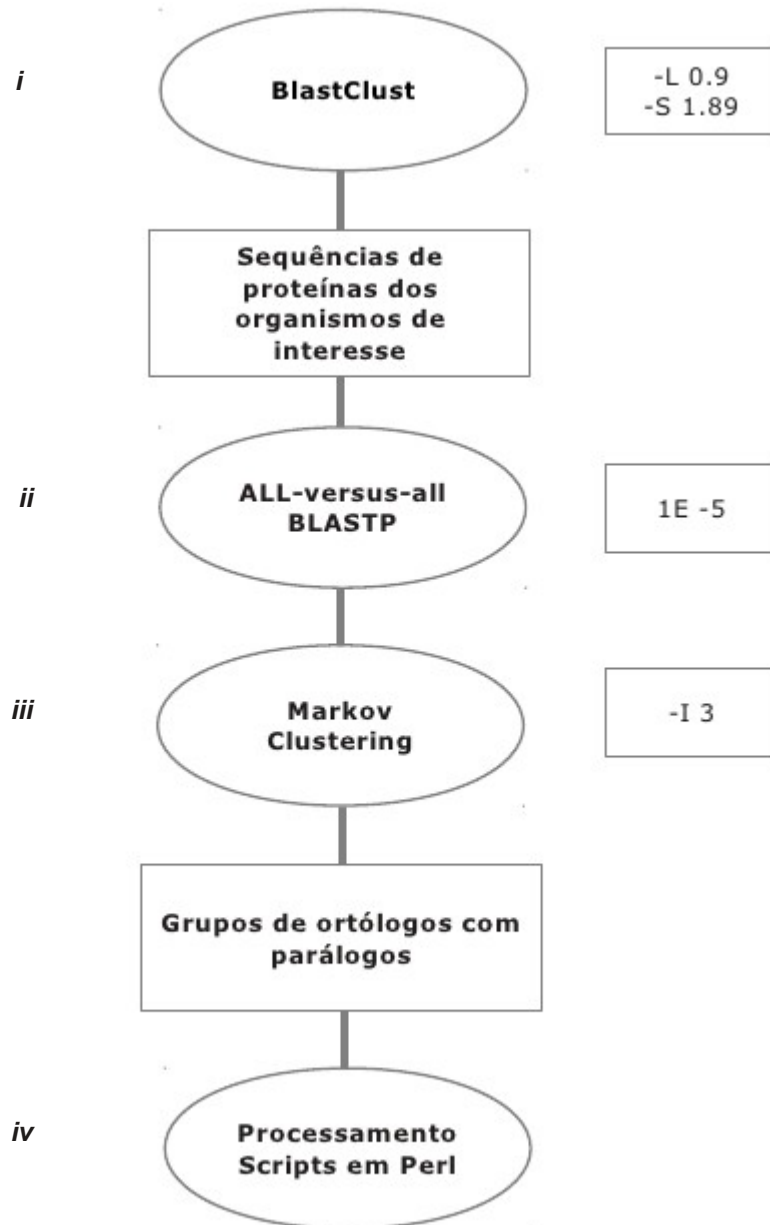


Figura 5: Fluxo gráfico do algoritmo MCL para o agrupamento de proteínas. (i) Primeiro foi utilizado BLASTCLUST para remover isoformas de splicing. (ii) Segundo passo foi BLASTP all-versus-all. (iii) Os resultados do BLASTP alimentaram o Markov Cluster Algorithm (MCL). (iv) Em seguida os clusters contendo as proteínas agrupadas por similaridade de sequência foram processados com scripts em Perl. Fonte: Adaptado de Li Li *et al.*, 2013.

4.6- Análises Estatísticas

As análises estatísticas foram conduzidas no ambiente estatístico R (<http://www.r-project.org/>). Foram utilizados Wilcoxon Test e Fisher Exact Test para calcular a significância estatística ($p < 0,05$) dos resultados apresentados.

5- Resultados

5.1- Agrupamento dos MDRs de *S. cerevisiae* e análise dos domínios protéicos

5.1.1- Muitos MDRs de *S. cerevisiae* surgiram por eventos de duplicação

Foram identificados 580 MDRs o que corresponde a 9,81% das proteínas do proteoma da *S. cerevisiae* que possui 5.911 proteínas preditas. Todas as proteínas dessa levedura foram agrupadas em famílias protéicas, de acordo com a similaridade entre as sequências, utilizando o algoritmo MCL. Após o agrupamento de todas as proteínas de *S. cerevisiae* em famílias foram encontradas 541 famílias que possuem pelo menos um MDR e outras 3.847 famílias que possuem genes não-MDRs. Foi observado que 65,86% dos MDRs (382 genes) são genes únicos (*i.e.* não foi detectado outro parálogo da levedura na mesma família) e 34,14% (198 genes) possuem parálogos próximos. Dentro do grupo dos MDRs com parálogos 64,64% (128) das famílias possuem parálogos não MDRs e 35,36% (70) possuem parálogos que também são MDRs (23 famílias possuem dois MDRs e oito famílias possuem três MDRs). As famílias protéicas sem MDRs possuem mediana de 1,29 parálogos, mais parálogos do que as famílias protéicas com MDRs (1,07) (Wilcoxon Test (WT); $p = 2,299 \times 10^{-11}$) (Figura 6).

Também foram utilizados os genes duplicados do trabalho de Fares *et al.*, 2013 para identificar os MDRs duplicados WGD e SSD e conseqüentemente os MDRs únicos (sMDRs). Foram encontrados 210 (36,21%) MDRs duplicados (dMDRs) sendo que 94 são MDRs-WGD e 82 MDRs-SSD, enquanto que 370 (63,79%) são sMDRs no genoma da levedura. Apesar da pequena diferença entre o número de sMDRs e dMDRs encontrados pelo presente trabalho no agrupamento protéico (382 e 198) e nos dados de Fares *et al.* 2013 (370 e 210) estes dados se corroboram entre si.

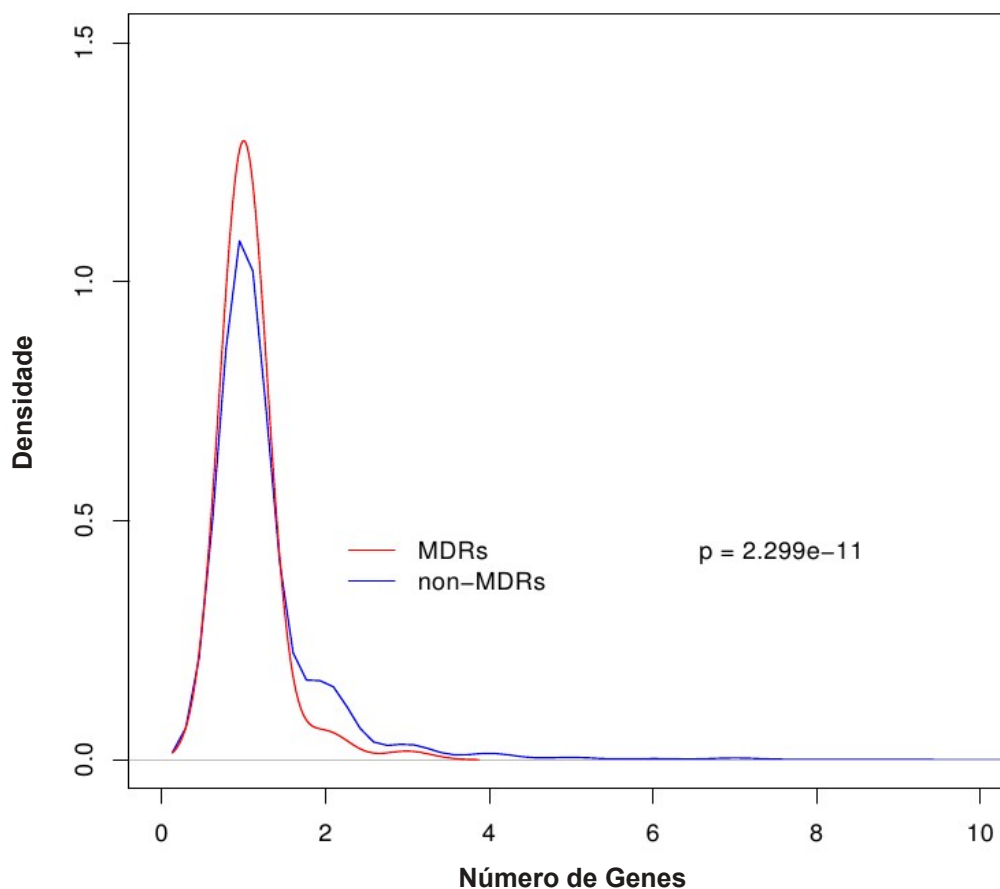
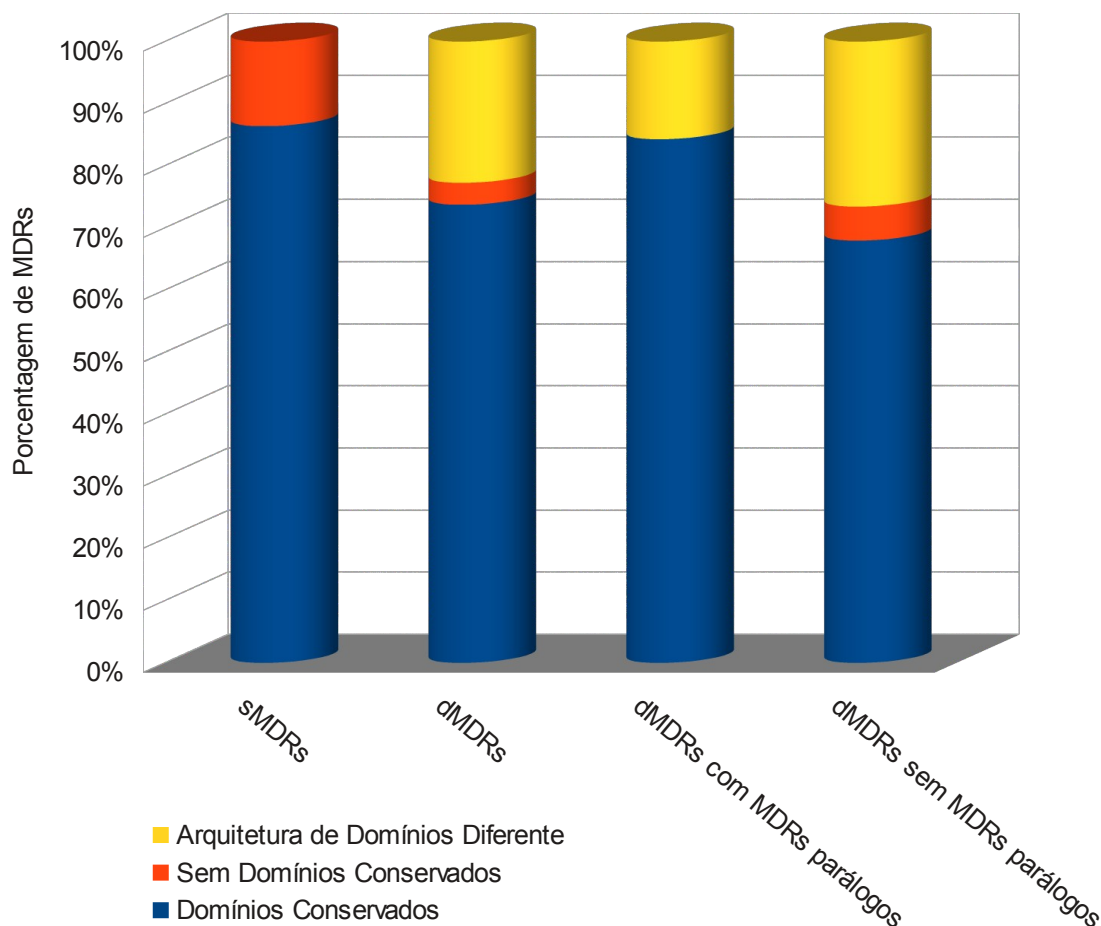


Figura 6: Distribuição dos genes em famílias proteicas com MDRs e sem MDRs (non-MDRs). A maioria das famílias proteicas com MDRs possuem apenas um gene enquanto que as famílias proteicas sem MDRs tendem a ter mais parálogos ($p = 2,299 \times 10^{-11}$). Densidade (Densidade Kernel) é uma forma não-paramétrica para estimar a função de densidade de probabilidade de uma variável aleatória.

5.1.2- A maioria dos MDRs de *S. cerevisiae* possuem domínios proteicos conservados idênticos aos seus parálogos

Para verificar se os MDRs possuem domínios conservados no Pfam e se dMDRs e seus respectivos parálogos possuem domínios idênticos utilizou-se HMMER e o banco de domínios conservados Pfam-A. Entre os sMDRs 86,39% (330) possuem domínios conservados e 13,61% (52) não possuem domínios conservados. Dentro do grupo dos dMDRs 73,73% (146) possuem domínios idênticos aos seus parálogos, 22,72% (45) possuem arquitetura de domínio diferente do seu parálogo e 3,53% (7) não possuem domínios conservados. No grupo dos dMDRs existem dois subgrupos: MDRs com parálogos MDRs e MDRs com

parálogos não MDRs. No primeiro grupo 84,28% (59) os MDRs possuem domínios idênticos aos seus MDRs parálogos e 15,71% (11) os MDRs possuem arquitetura de domínio diferente. No segundo grupo 67,97% (87) possuem domínios idênticos aos seus parálogos não MDRs, 26,56% possuem arquitetura de domínio diferente e 5,47% não possuem domínios conservados (Figura 7).



*Figura 7: Análise dos domínios conservados dos MDRs de *S. cerevisiae*. Em azul estão representados os MDRs com domínios conservados e/ou idênticos. Os MDRs sem domínios conservados estão representados em vermelho. dMDRs com arquitetura de domínios diferente estão representados em amarelo.*

5.2- Análise do surgimento de mutações em MDRs através das gerações em *S. cerevisiae*

Os resultados apresentados a seguir foram realizados com dados do sequenciamento do genoma completo de várias populações de *S. cerevisiae*, que identificaram mutações em 1000 gerações de 40 populações de levedura crescendo em meio rico (Lang *et al.*, 2013) (Mais informações no item 4.4).

5.2.1- Os MDRs são proporcionalmente mais representados entre os “putative drivers” do que os não MDRs

Lang *et al.* 2013 selecionaram 24 genes com três ou mais mutações e os chamaram de “putative drivers”, genes supostamente envolvidos em processos adaptativos, que quando mutados, seriam benéficos para a população de *S. cerevisiae*. Proporcionalmente existem mais MDRs entre os “putative drivers” do que não MDRs, 7 genes representando 1.21% dos MDRs e 17 genes os não MDRs (0.32%) (Fisher Exact Test (FET); $p = 0.006608$). Os MDRs “putative drivers” são *ROT2*, *PDR5*, *IRA2*, *ANP1*, *CNE1*, *GAS1* e *GPB1*. Eles estão envolvidos em processos biológicos como: biogêneses da parede celular, transporte de multidrogas, resposta a estresse, regulador negativo de Ras, glicosilação de proteína, enovelamento de proteínas e montagem da parede celular (Tabela).

Tabela: Os genes mais mutados são os “putative drivers” da adaptação

GI	ORF	Gene	Hits	Fixada	Processos Biológicos	Ortologia	Classificação
398365033	YBR140C	<i>IRA1</i>	21	10	Regulador Negativo de Ras	Percolozoa	PAR-WGD
398365611	YBR229C	<i>ROT2</i>	11	2	Biogêneses da Parede Celular	Percolozoa	sMDR
6322322	YJL139C	<i>YUR1</i>	11	5	Biogêneses da Parede Celular	Microsporidia	Não MDR
6323160	YLR131C	<i>ACE2</i>	9	4	Citocinese	Neocallimastigomycota	Não MDR
6323394	YLR362W	<i>STE11</i>	9	1	Cruzamento	Percolozoa	PAR-SSD
6321876	YHR084W	<i>STE12</i>	9	2	Cruzamento	Microsporidia	Não MDR
398365429	YOR153W	<i>PDR5</i>	8	5	Transporte de Multidrogas	Percolozoa	MDR-WGD
6324617	YOR043W	<i>WHI2</i>	7	2	Resposta a Estresse Geral	Saccharomyces	Não MDR
398365783	YOR212W	<i>STE4</i>	6	1	Cruzamento	Percolozoa	Não MDR
6324491	YOL081W	<i>IRA2</i>	5	3	Regulador Negativo de Ras	Percolozoa	MDR-WGD
6325417	YPR159W	<i>KRE6</i>	4	1	Montagem da Parede Celular	Kickxellomycotina	Não MDR
398365337	YOR140W	<i>SFL1</i>	4	1	Regulação dos genes de Floculação	Pezizomycotina	Não MDR
6320308	YDR103W	<i>STE5</i>	4	3	Cruzamento	Saccharomyces	Não MDR
6320799	YEL036C	<i>ANP1</i>	3	1	Glicosilação de Proteínas	Chytridiomycota	MDR-SSD
6319260	YAL058W	<i>CNE1</i>	3	2	Enovelamento de Proteínas	Percolozoa	sMDR
6323967	YMR307W	<i>GAS1</i>	3	3	Montagem da Parede Celular	Kickxellomycotina	MDR-SSD
398364287	YGL195W	<i>GCN1</i>	3	1	Regulação da Tradução	Percolozoa	Não MDR
398366409	YOR371C	<i>GPB1</i>	3	1	Regulador Negativo de Ras	Saccharomyces	MDR-WGD
330443366	YAL056W	<i>GPB2</i>	3	1	Regulador Negativo de Ras	Saccharomyces	PAR-WGD
14318567	YFR042W	<i>KEG1</i>	3	0	Montagem da Parede Celular	Saccharomyces	Não MDR
6324912	YOR336W	<i>KRE5</i>	3	1	Montagem da Parede Celular	Saccharomyces	Não MDR
6324690	YOR116C	<i>RPO31</i>	3	0	Transcrição da RNA polimerase III	Percolozoa	Não MDR
6322356	YJL105W	<i>SET4</i>	3	2	Desconhecido	Chytridiomycota	PAR-WGD
330443603	YJL171C	<i>YJL171C</i>	3	0	Desconhecido	Candida	Não MDR

GI = Identificação do Gene, ORF = Open Reading Frame, Gene = Nome do Gene, Hits = Número de Mutações, Fixada = Número de Mutações Fixadas
 PAR-WGD = Parálogo WGD de MDR, PAR-SSD = Parálogo SSD de MDR, MDRs = Negrito

5.2.2- Os MDRs sofrem mais mutações que os não MDRs

Existem 580 MDRs e 5.331 não MDRs no genoma de *S. cerevisiae* e foram identificados 71 MDRs e 475 não MDRs mutados correspondendo a 12,24% e 8,97% respectivamente (FET; $p = 0,01$). Os MDRs são portanto proporcionalmente mais mutados do que os não MDRs. Os MDRs possuem mediana de 1,62 mutações por gene enquanto que os não MDRs possuem somente 1,27, a diferença é estatisticamente significativa (WT; $p = 0,011$). A mediana da frequência de mutação dos MDRs através das gerações cresce mais do que a dos não MDRs (WT; $p = 0,009$) (Figura 8). A mediana da fixação de mutação também é maior nos MDRs (0,509) do que nos não MDRs (0,265) (WT; $p = 0,012$).

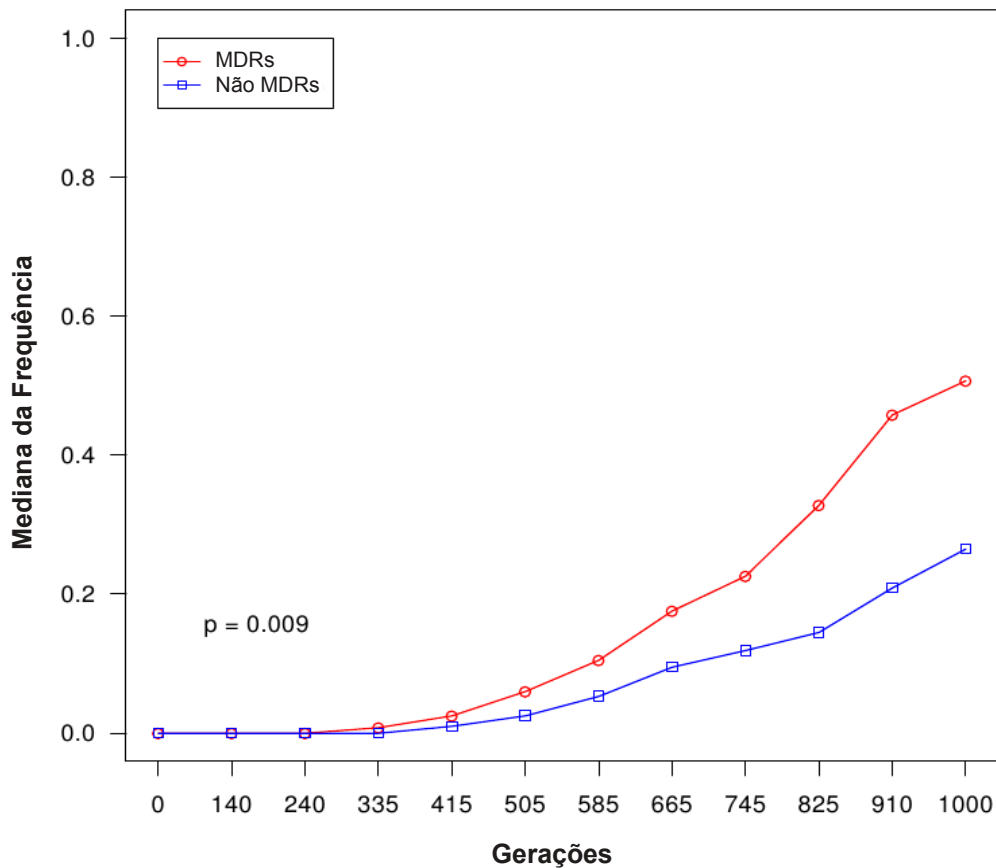


Figura 8: *Frequência das mutações ao longo das gerações de S. cerevisiae. Os MDRs possuem maior frequência de mutação do que os não MDRs ($p = 0,009$).*

5.2.3- A frequência de mutação dos dMDRs é maior do que os sMDRs

Existem 210 dMDRs e 370 sMDRs. Foram encontrados 31 dMDRs e 40 sMDRs com mutações, correspondendo respectivamente a 14,76% e 10,81% do total de genes de cada grupo (FET; $p = 0,187$). A mediana da taxa de mutação por gene é similar entre dMDRs (1,61) e sMDRs (1,60) (WT; $p = 1,0$). Mas a mediana da frequência de mutação é significativamente maior nos dMDRs do que nos sMDRs (WT; $p = 0,009$), sugerindo que genes com parálogos próximos apresentam relaxamento na pressão seletiva. (Figura 9). A mediana da taxa de fixação das mutações dos dMDRs (0,731) é maior do que a dos sMDRs (0,369), mas esta diferença não é estatisticamente significativa (WT; $p = 0,075$). A Tabela mostra que existem apenas dois sMDRs entre os “putative drivers” e ambos precedem ao SUI

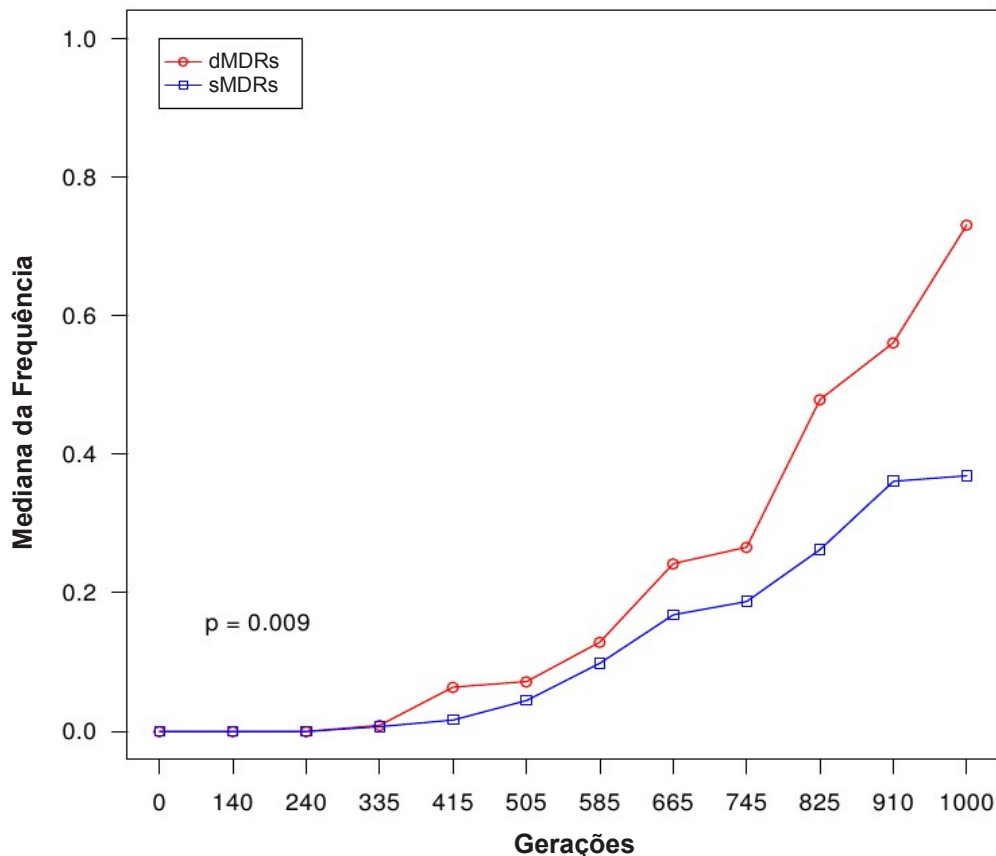


Figura 9: *Frequência das mutações ao longo das gerações de S. cerevisiae. Os dMDRs possuem maior frequência de mutações que os sMDRs ($p = 0,009$).*

5.2.4- Os MDRs sofrem menos mutações que os seus parálogos

Existem 210 pares de dMDRs e seus parálogos, eles possuem o mesmo número de genes mutados (29 cada), mas a mutação não é em ambos genes do par. A mediana da taxa de mutação é diferente em cada grupo. Os parálogos dos dMDRs possuem 2,21 mutações por gene enquanto os dMDRs somente 1,65 (WT; $p = 7,177 \times 10^{-07}$). A mediana da frequência de mutação é significativamente maior nos dMDRs do que nos seus parálogos (WT; $p = 0,009$) (Figura 10). A mediana da fixação da mutação também é maior nos dMDRs (0,718) do que nos seus parálogos (0,406), mas não é estatisticamente significativa (WT; $p = 0,203$).

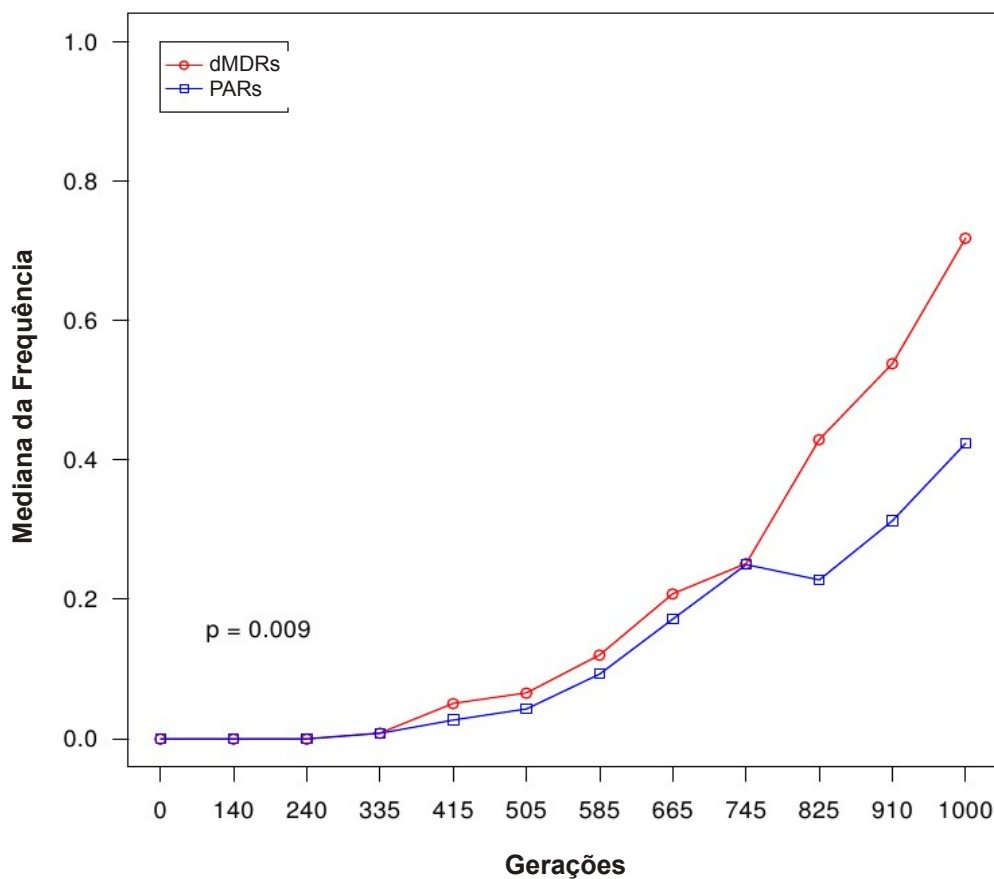


Figura 10: *Frequência das mutações ao longo das gerações de *S. cerevisiae*. Os dMDRs possuem maior frequência de mutação do que os seus parálogos (PARs ($p = 0,009$)).*

5.2.5- As mutações através das gerações são similares entre MDRs-WGD e seus parálogos WGD

Existem 94 pares de MDRs-WGD e parálogos. Foi identificado uma proporção semelhante de MDRs-WGD (17,35%) e de parálogos WGD (12,24%) mutados (FET; $p = 0,421$). Os parálogos WGD possuem uma mediana da taxa de mutação de 3,25, sendo maior que a dos MDRs-WGD (1,94), mas não é estatisticamente significativa (WT; $p = 0,346$). A mediana da frequência de mutação é semelhante em ambos os grupos (WT; $p = 0,221$) (Figura 11). A mediana da taxa de fixação da mutação também é semelhante entre os MDRs-WGD (0,623) e os seu parálogos WGD (0,675) (WT; $p = 0,871$). Apesar da semelhança mutacional entre MDRs-WGD e seus parálogos WGD o presente trabalho identificou que o gene mais mutado nos experimentos de Lang *et al.*, 2013 é um parálogo WGD de um MDR (Tabela). Este gene é *IRA1* um regulador negativo de RAS com 21 mutações e 10 fixadas. Lang *et al.*, 2013 sugerem que a seleção neste gene é para perda de função porque foram observadas 14 mutações *missense*, 8 *nonsense* e 10 *frameshift*.

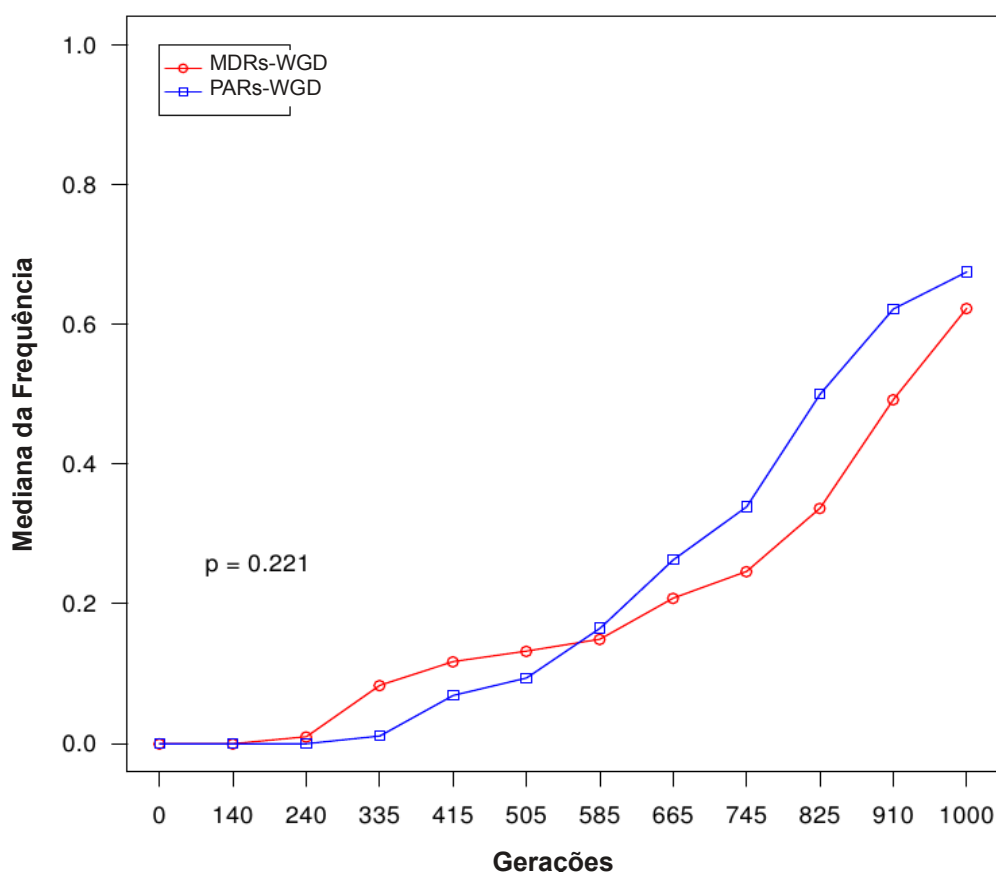


Figura 11: **Frequência das mutações ao longo das gerações de *S. cerevisiae*.** Os MDRs-WGD possuem frequência de mutação semelhante aos seus parálogos WGD (PARs-WGD) ($p = 0,221$).

5.2.6- OS MDRs-SSD com parálogos SSD possuem maior frequência de mutação

Existem 82 pares de MDRs-SSD e parálogos. A pressão seletiva age de forma diferente entre MDRs-SSD e seus parálogos SSD. Especificamente, 15,17% dos parálogos SSD e 10,71% dos MDRs-SSD foram mutados (WT; $p = 0,426$). A mediana da taxa de mutação é similar entre os parálogos SSD (1,47) e os MDRs-SSD (1,33) (WT; $p = 0,426$). Mas a mediana da frequência de mutações é maior nos MDRs-SSD do que nos parálogos SSD (WT; $p = 0,014$) (Figura 12). A mediana da taxa de fixação da mutação também é maior nos MDRs-SSD (0,887) do que nos seus parálogos SSD (0,217) (WT; $p = 0,0267$). Estes resultados sugerem que os MDRs-SSD quando mutados apresentam potencial adaptativo. Existem dois MDRs-SSD entre os “putative drivers” (*ANP1* e *GAS1*), ambos surgiram no ancestral comum dos taxa Chytridiomyta e Kickxellomycota respectivamente (Tabela).

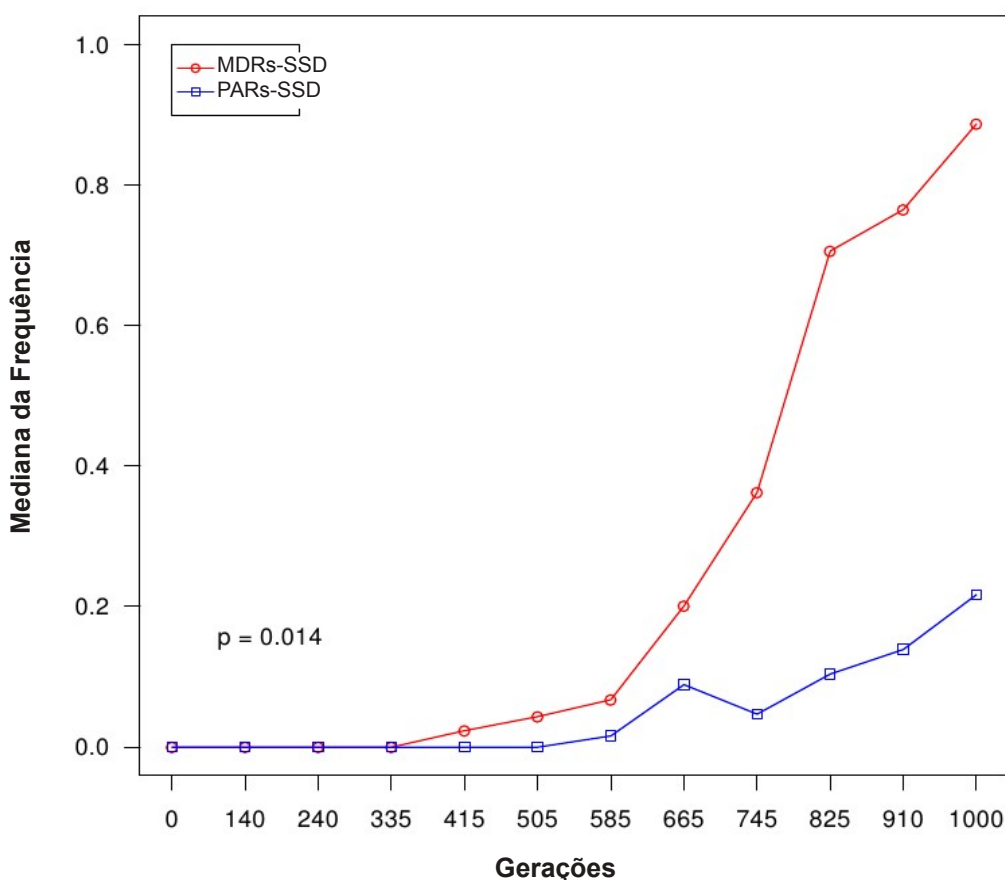


Figura 12: *Frequência das mutações ao longo das gerações de S. cerevisiae. Os MDRs-SSD possuem maior frequência de mutação do que seus parálogos SSD (PARs-SSD) ($p = 0,014$).*

5.2.7- Os MDRs-WGD e os MDRs-SSD possuem frequência de mutação similar

Existem 94 MDRs-WGD e 82 MDRs-SSD. Embora sem atingir a significância estatística há mais MDRs-WGD (17,35%) mutados do que MDRs-SSD (10,71%) (FET; $p = 0,229$). A mediana da taxa de mutação é similar entre os MDRs-WGD (1,94) e os MDRs-SSD (1,33) (WT; $p = 0,460$). A mediana da frequência de mutação é semelhante entre os grupos (WT; 0,749) (Figura 13). A mediana da taxa de fixação da mutação é maior nos MDRs-SSD (0,887) do que nos MDRs-WGD (0,623), mas este resultado não atingiu significância estatística (WT; $p = 0,445$). A frequência de mutação é similar entre os grupos analisados sugerindo que a origem dos MDRs (WGD ou SSD) não interfere no padrão mutacional.

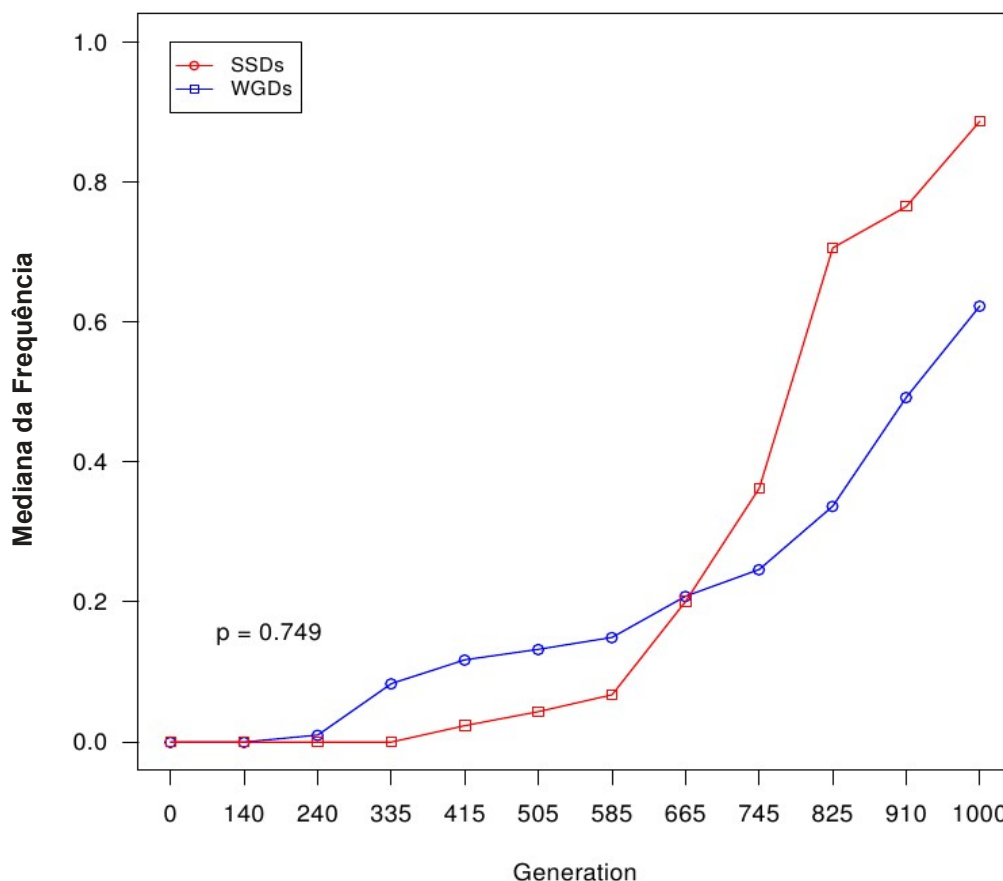


Figura 13: *Frequência das mutações ao longo das gerações de S. cerevisiae. Os MDRs-SSD possuem frequência de mutação similar aos MDRs-WGD ($p = 0,749$).*

5.3- Análise do perfil funcional dos MDRs de *S. cerevisiae*

Os resultados apresentados a seguir foram obtidos a partir de análise de dados de interações genéticas (GIs) e interações proteína-proteínas (PPIs) (Mais informações no item 4.3).

5.3.1- Os MDRs possuem maior diversidade funcional do que os não MDRs

Existem 580 MDRs e 5.331 não MDRs no genoma de *S. cerevisiae*. Foram identificados 482 MDRs e 3.462 não MDRs com informações de GIs. Analisando essas GIs foi possível observar que os MDRs possuem mediana de 117,5 GIs, que é aproximadamente três vezes maior que os não MDRs com mediana de 36 GIs (WT; $p < 2,2 \times 10^{-16}$) (Figura 14 A). Os dados de PPIs corroboram com os resultados apresentados acima, os MDRs (576) possuem mediana de 20 PPIs enquanto os não MDRs (4.911) possuem somente 13 (WT; $p < 2,2 \times 10^{-16}$) (Figura 14 B). Baseado nestes resultados é possível concluir que os MDRs possuem maior diversidade funcional que os não MDRs.

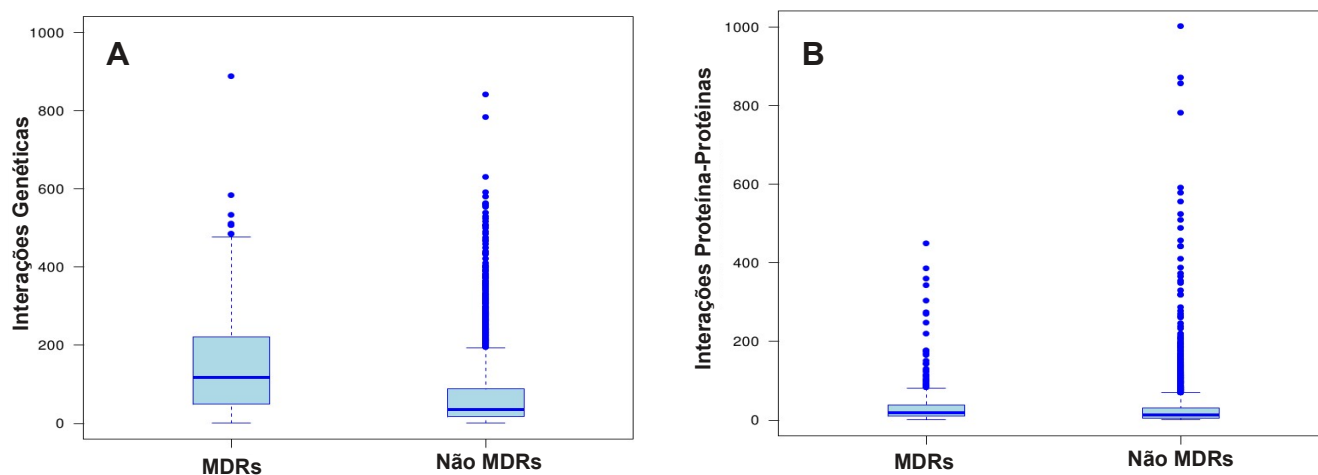


Figura 14: Gráfico box-plot das interações genéticas e proteína-proteínas. MDRs versus não MDRs.

5.3.2- Os dMDRs e sMDRs possuem número de GIs e PPIs similar

Existem 210 dMDRs e 370 sMDRs e foram identificados 182 dMDRs e 300 sMDRs com informações de GIs. Analisando esses dados observou-se que os dMDRs possuem mediana de 105 GIs, enquanto os sMDRs possuem mediana de 125,5 GIs (WT; $p = 0,028$) (Figura 15 A). Dentro dos dMDRs 207 genes possuem mediana de 25 PPIs, enquanto 369 sMDRs possuem mediana de 19 PPIs (WT; $p = 0,086$) (Figura 15 B). Os resultados demonstram que os dMDRs e os sMDRs possuem número de GIs e PPIs similar sugerindo que o evento de duplicação não interfere no destino funcional dos MDRs.

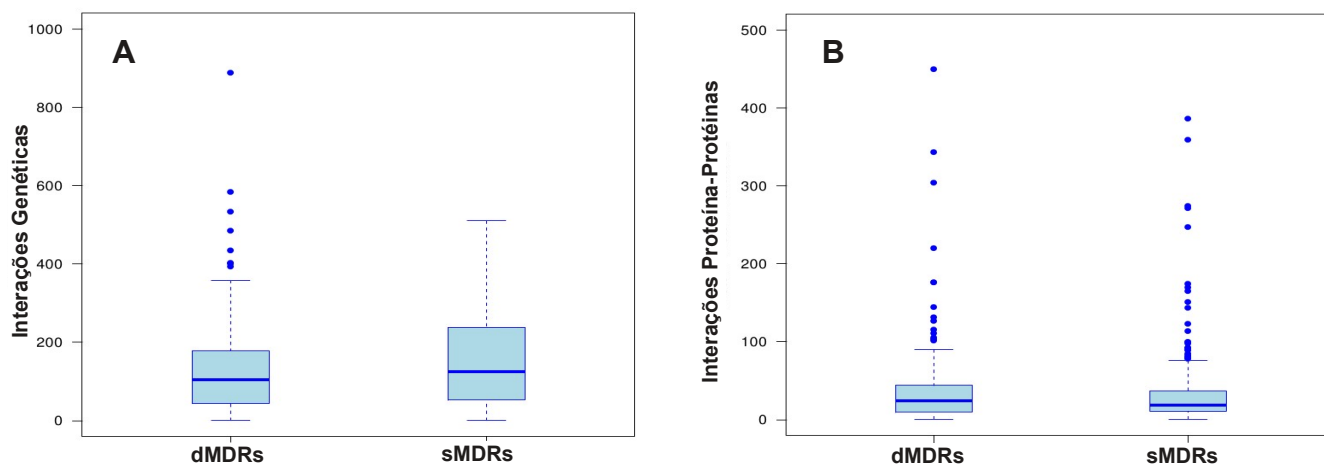


Figura 15: Análise das interações genéticas e proteína-proteínas. dMDRs versus sMDRs.

5.3.3- Os dMDRs possuem maior diversidade funcional que os seus parálogos

Existem 210 pares de dMDRs com parálogos e foram identificados 132 dMDRs com informações de GIs, que possuem uma mediana de 100 GIs enquanto os seus 132 parálogos possuem mediana de 32,5 GIs (WT; $p = 6,85 \times 10^{-7}$). A razão entre dMDRs e seus parálogos permite confirmar se os dMDRs de cada par possui mais GIs ou PPIs que seus parálogos. A mediana da razão (escala logarítmica) de GIs entre dMDRs e parálogos é de 1,28 (WT; $p = 1,71 \times 10^{-8}$) (Figura 16 A), corroborando com o resultado anterior. Os 169 dMDRs com informações de PPIs

possuem uma mediana de 26 PPIs e os 169 parálogos 14 PPIs (WT; $p = 1,407 \times 10^{-8}$). Consequentemente a mediana da razão de PPIs é significativamente maior que zero (0,57) (WT; $3,662 \times 10^{-8}$) (Figura 16 B). Estes resultados permitem concluir que os MDRs possuem maior diversidade funcional que os seus parálogos.

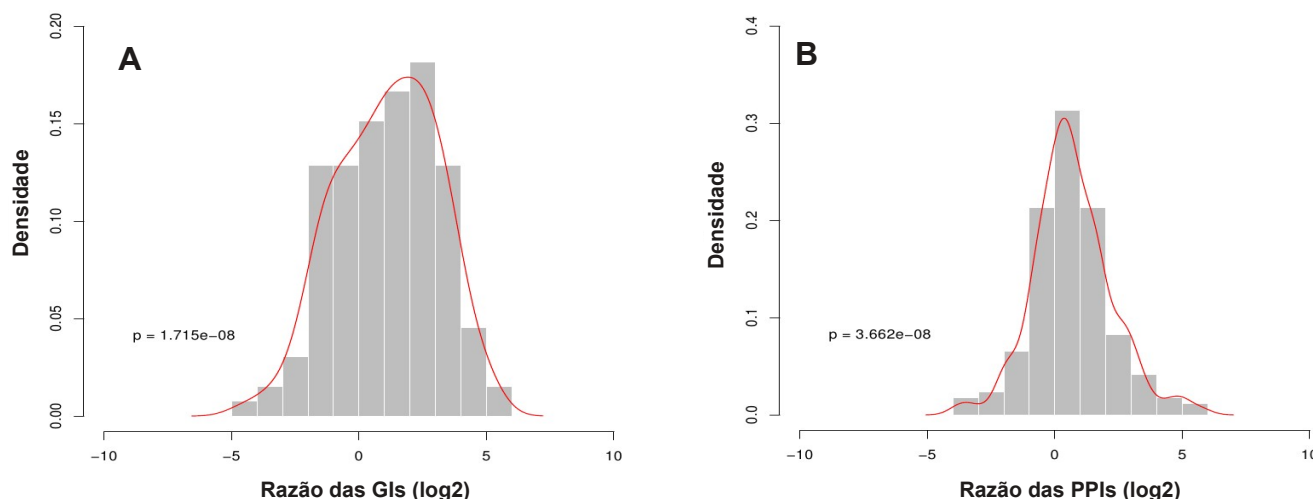


Figura 16: Razão das interações genéticas e protéicas dos MDRs e seus parálogos. MDRs versus Parálogos. Densidade (Densidade Kernel) é uma forma não-paramétrica para estimar a função de densidade de probabilidade de uma variável aleatória.

5.3.4- Os parálogos WGD dos MRDs-WGD estão perdendo função após a duplicação

Existem 94 pares de MDRs-WGD com parálogos e foram identificados 76 MDRs-WGD com informações de GIs, que possuem mediana de 108 GIs enquanto 76 parálogos dos MDRs-WGD possuem mediana de 25,5 GIs (WT; $p = 6,581 \times 10^{-9}$). A razão entre MDRs-WGD e seus parálogos permite confirmar se os MDRs-WGD de cada par possui mais GIs ou PPIs que seus parálogos. A mediana da razão (escala logaritmica) é 1,85 (WT; $p = 1,386 \times 10^{-8}$) (Figura 17 A), corroborando com o resultado anterior. Existem 89 pares de MDRs-WGD e parálogos com informações de PPIs. A mediana de PPIs é duas vezes maior nos MDRs-WGD (27) do que nos seus parálogos WGD (13) (WT; $2,82 \times 10^{-8}$). Resultado semelhante foi encontrado ao analisar a mediana da razão de PPIs (0,75) (WT; $p = 2,84 \times 10^{-8}$) (Figura 17 B). Foram encontrados apenas dois pares de parálogos MDRs-WGD onde ambos são MDRs: *VPS5* e *YKR078W*; *SIP3* e *YSP1*.

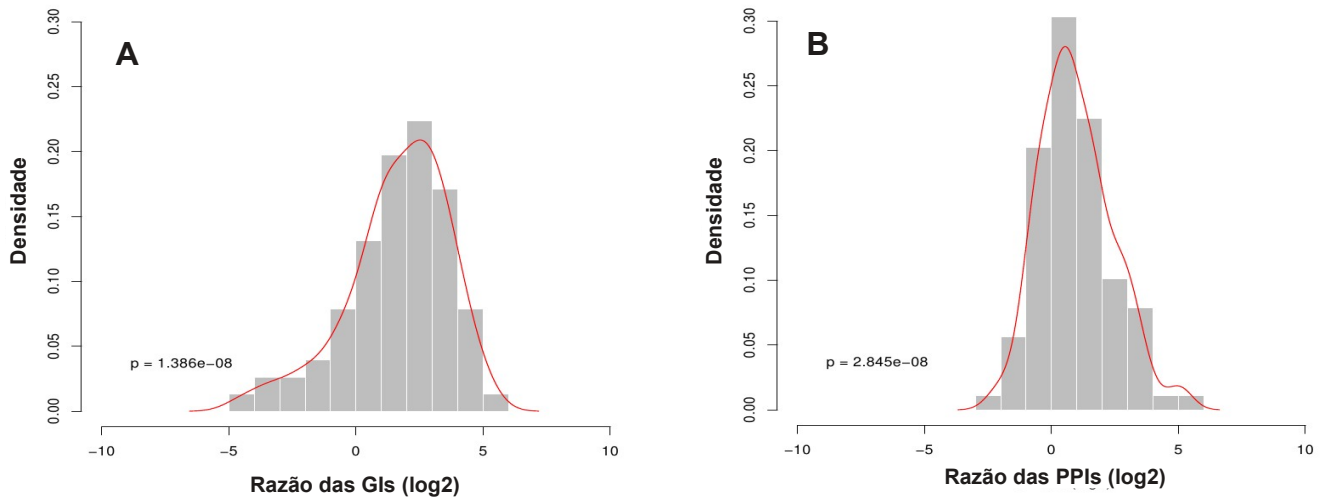


Figura 17: Razão das interações genéticas e protéicas dos MDRs-WGD e seus parálogos. MDRs-WGD versus parálogos WGD. Densidade (Densidade Kernel) é uma forma não-paramétrica para estimar a função de densidade de probabilidade de uma variável aleatória.

5.3.5- Os MDRs-SSD e seus parálogos SSD estão ganhando novas funções após a duplicação

Existem 82 pares de MDRs-SSD com parálogos e foram identificados 56 pares com informações de GIs. Os MDRs-SSD possuem mediana de 91,5 GIs e os seus parálogos SSD 54 (WT; $p = 0,180$). A razão entre MDRs-SSD e seus parálogos permite confirmar se os MDRs-SSD de cada par possui mais GIs ou PPIs que seus parálogos. O resultado anterior corrobora com a mediana da razão (escala logarítmica) das GIs (0,11) entre MDRs-SSD e seus parálogos (WT; $p = 0,123$) (Figura 18 A). O número de PPIs é diferente entre os 80 pares de MDRs-SSD e seus parálogos. Os MDRs-SSD possuem mediana de 25,5 PPIs e os seus parálogos mediana de 17 PPIs (WT; $p = 0,018$). Conseqüentemente a razão das PPIs é significativamente maior que zero (0,24) (WT; $p = 0,018$) (Figura 18 B). Os resultados demonstram que os MDRs-SSD possuem número semelhante de GIs aos seus parálogos SSD, enquanto que o primeiro grupo possui mais PPIs que o segundo. Foram encontrados 15 pares de parálogos MDRs-SSD (ambos MDRs): *RRD1* e *RRD2*; *CKA1* e *CKA2*; *HDA2* e *HDA3*; *PEP12* e *TLG2*; *FBP26* e *PFK26*; *ERG4* e *ERG24*; *DID4* e *VPS24*; *VMA3* e *VMA11*; *SWI3* e *RSC8*; *ERG5* e *ERG11*; *MUP1* e *MUP3*; *CKB1* e *CKB2*; *BRO1* e *RIM20*; *ALG6* e *ALG8*; *NAB6* e *MRN1*.

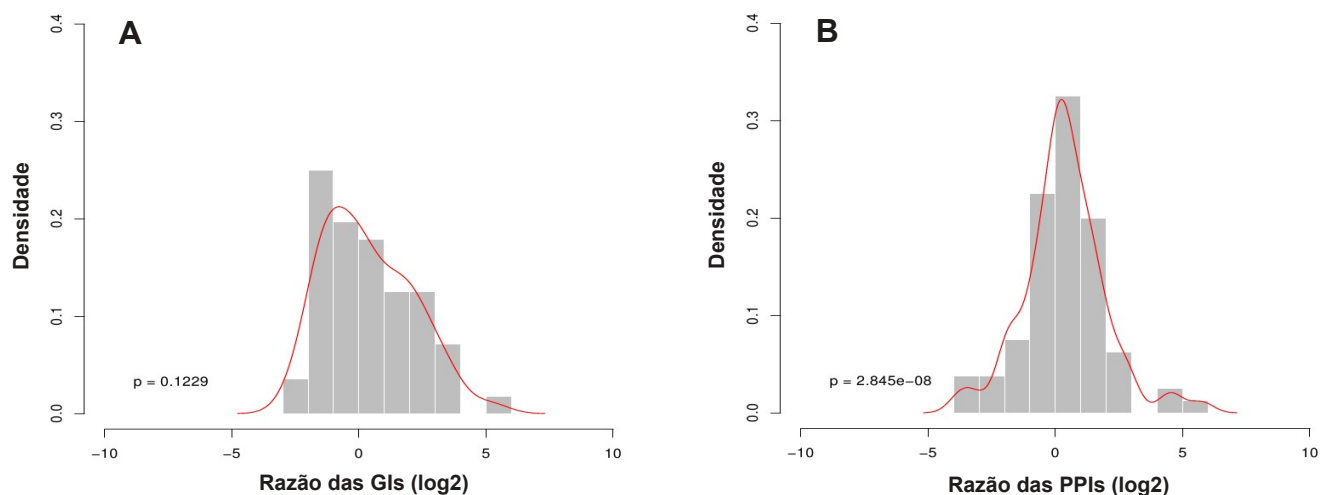


Figura 18: Razão das interações genéticas e protéicas dos MDRs e seus parálogos. MDRs versus parálogos SSD. Densidade (Densidade Kernel) é uma forma não-paramétrica para estimar a função de densidade de probabilidade de uma variável aleatória.

5.3.6- Os MDRs-WGD e os MDRs-SSD possuem números similares de interações protéica e genética

Existem 94 MDRs-WGD e 82 MDRs-SSD. Os 76 MDRs-WGD e os 56 MDRs-SSD com informações de GIs possuem medianas similares dessas interações. Os MDRs-SSD possuem 111,5 GIs e os MDRs-WGD 96 (WT; $p = 0,703$) (Figura 19 A). Os 89 MDRs-WGD e os 80 MDRs-SSD com informações de PPIs possuem medianas similares dessas interações. Os MDRs-SSD possuem 25,5 PPIs e os MDRs-WGD 26,5 (WT; $p = 0,715$) (Figure 19 B). Foi possível observar que os MDRs-WGD e os MDRs-SSD possuem medianas de GIs e PPIs semelhantes.

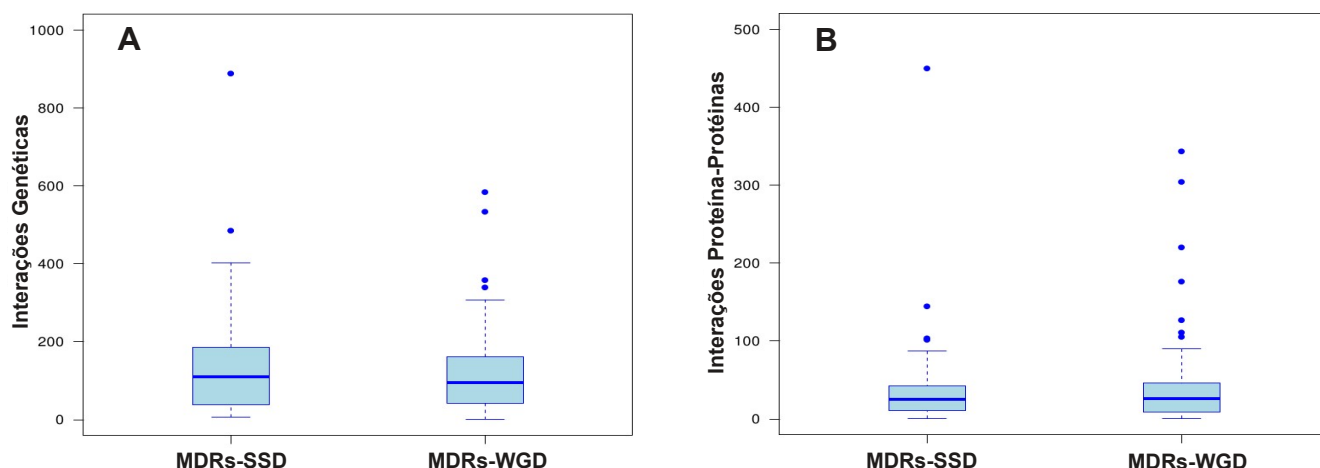


Figura 19: Análise das interações genéticas e proteína-proteínas. MDRs-SSD versus MDRs-WGD.

5.4- Análise da classificação taxonômica dos MDRs

Os resultados apresentados asseguir foram obtidos através da seleção de 40 proteomas de fungos, dividido em 13 taxa, e três proteomas protistas, que são externo ao filo em estudo. O agrupamento das proteínas em famílias foi realizado com o algoritmo MCL (Mais informações no item 4.5).

5.4.1- Os MDRs são mais ancestrais do que os não MDRs

Utilizou-se informações de famílias protéicas para estimar o surgimento dos MDRs através da evolução dos fungos. Por exemplo, se uma dada família protéica possui ortólogos dos taxa *Saccharomyces*, *Candida*, *Saccharomycotina*, *Pezizomycotina*, *Taphrinomycotina* e *Basidiomycota*, infere-se que esta família tenha surgido no ancestral comum entre o filo *Basidiomycota* e os demais taxa descritos acima, porque este é o táxon mais ancestral nesta família protéica. Ao aplicar sistematicamente este raciocínio, foi demonstrado que famílias com MDRs são mais ancestrais ao filo *Fungi* que famílias sem MDRs. Especificamente, 40,66% das famílias com MDRs precedem o surgimento dos fungos (versus 37,65% das famílias sem MDRs). Reciprocamente, encontramos que somente 25,69% das famílias com MDRs são específicas do clado *Saccharomyces* (contra 32,66% das famílias sem MDRs) (Figura 20).

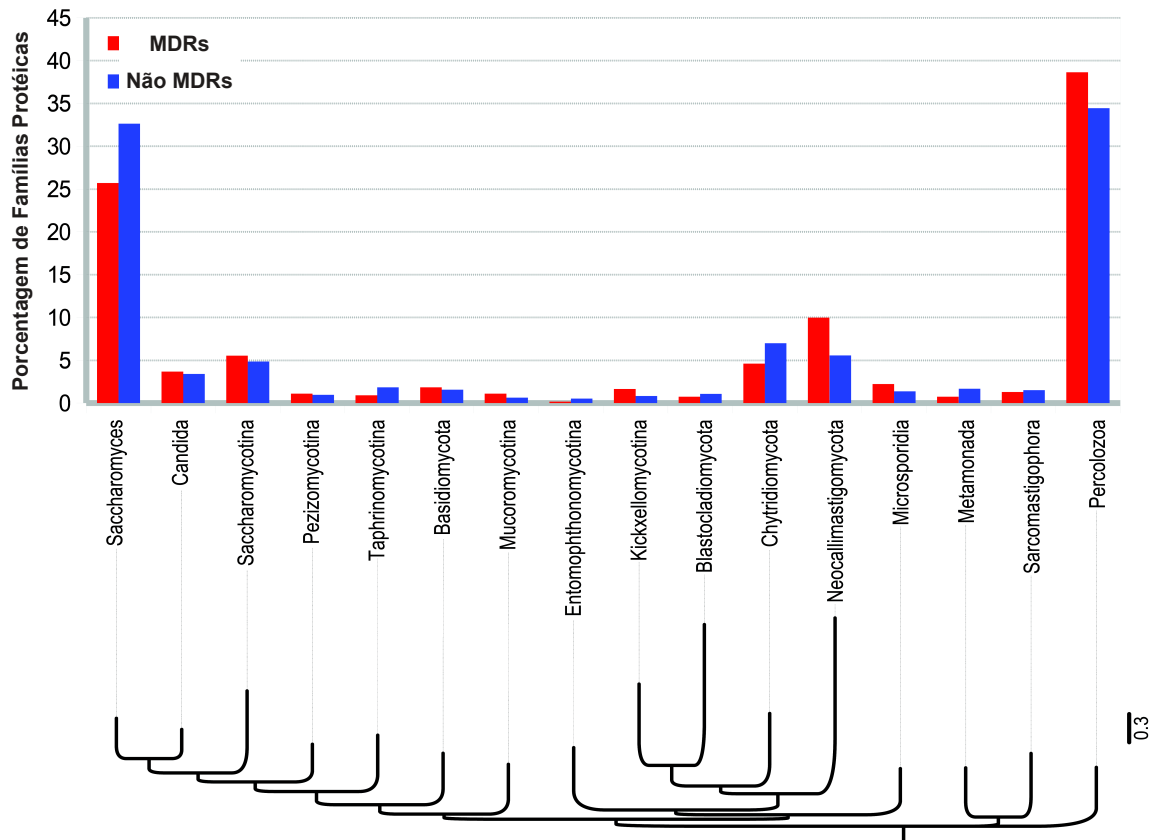


Figura 20: *Classificação taxonômica dos MDRs.* A figura mostra a porcentagem de clusters por taxon com MDRs e sem MDRs. A filogenia por maximum-likelihood da proteína ribossomal 40 S demonstra da esquerda para direita os 13 primeiros taxa pertencente ao filo Fungi e os três últimos pertencente ao filo Protista. O valor do Bootstrap está representado na barra a direita.

5.4.2- Os dMDRs são muito mais conservados do que os sMDRs

Foi possível observar que 59,18% das famílias protéicas com dMDRs precedem o surgimento dos fungos (versus 39,58% das famílias com sMDRs). Alternativamente, foram encontrados 32,88% das famílias com sMDRs surgindo no ancestral comum do clado Saccharomyces (contra 10,34% das famílias dos dMDRs) (Figura 21).

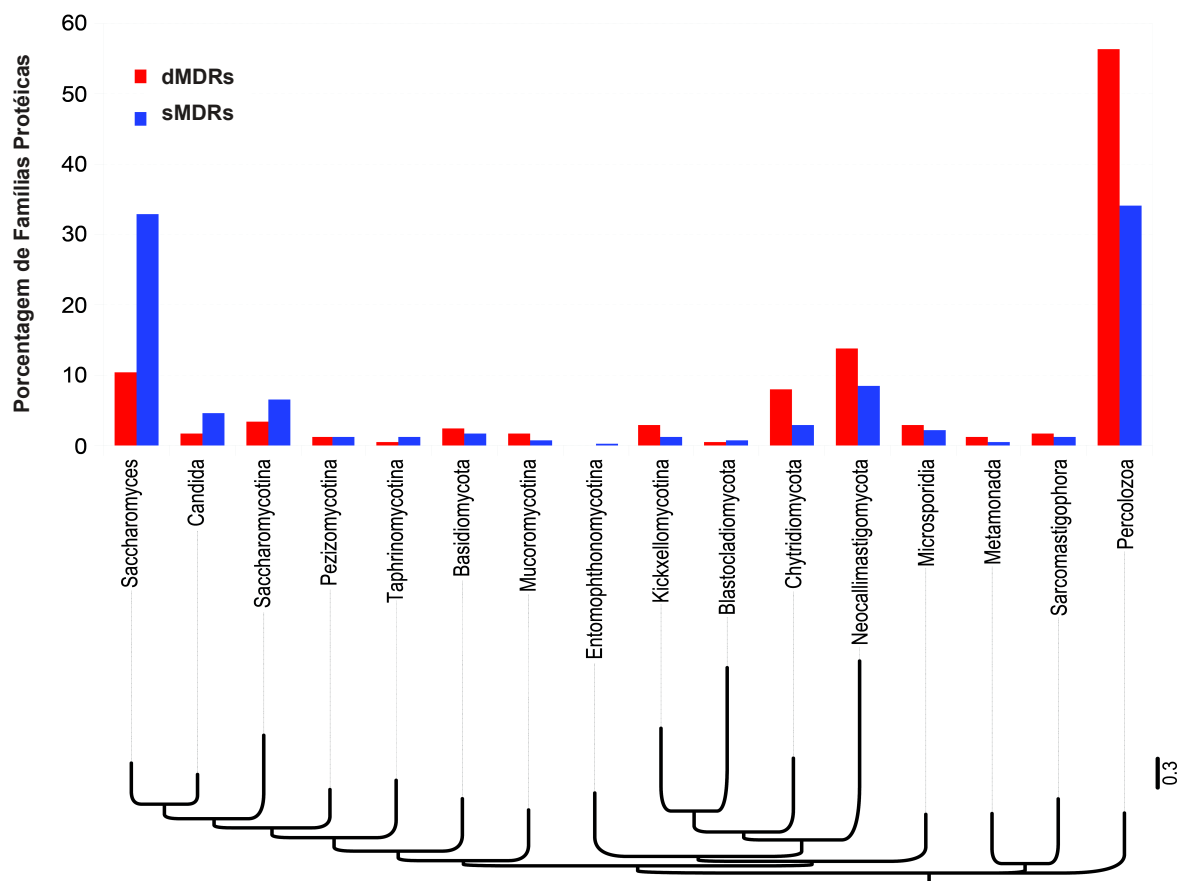


Figura 21: **Classificação taxonômica dos MDRs.** A figura mostra a porcentagem de famílias protéicas por taxon com dMDRs e sMDRs. Idem Figura 20.

5.4.3- Os dMDRs são mais conservados que seus parálogos

As famílias protéicas com MDRs tendem a ser mais conservadas do que as famílias com seus parálogos (Figura 22). Identificou-se que 54,64% das famílias com MDRs precedem o surgimento dos fungos (contra 50% das famílias com seus parálogos). Correlativamente, 12,5% das famílias com MDRs surgiram no ancestral comum do taxon *Saccharomyces* (versus 9,28% das famílias com seus parálogos).

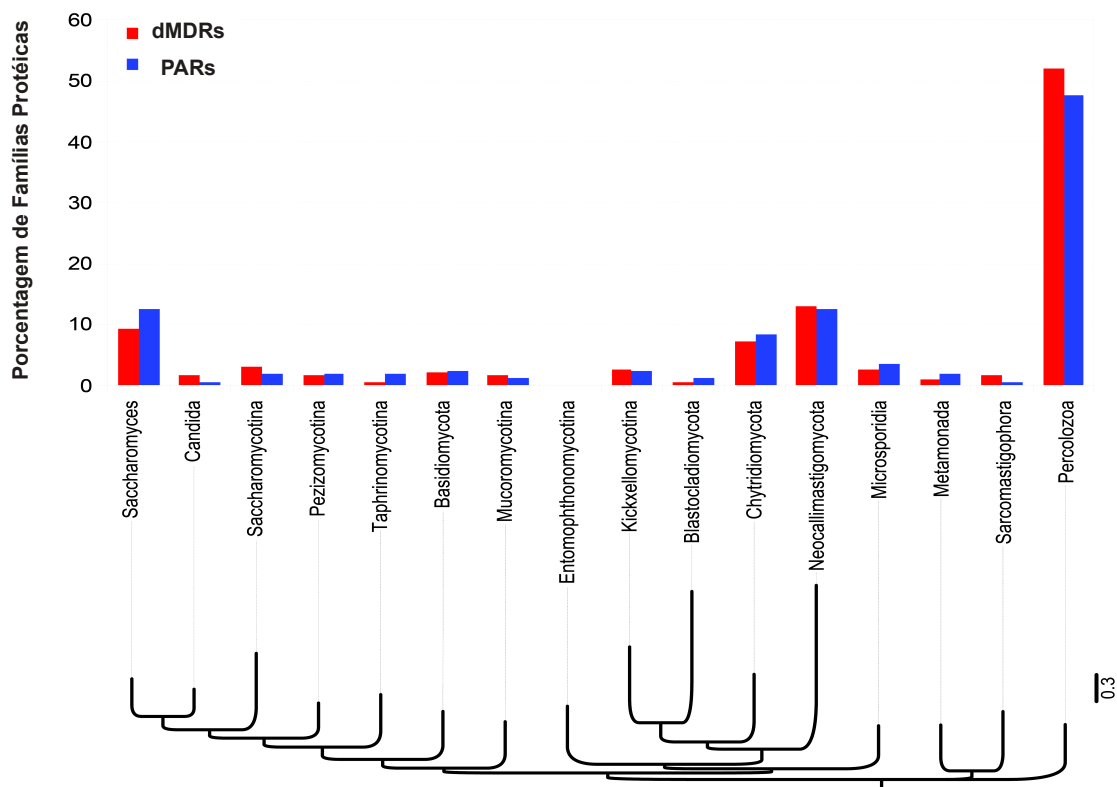


Figura 22: **Classificação taxonômica dos MDRs.** A figura mostra a porcentagem de famílias protéicas por taxon com MDRs e parálogos (PARs). Idem Figura 20.

5.4.4- Os MDRs-WGD e seus parálogos WGD possuem conservação similar através dos taxa

As famílias protéicas dos MDRs-WGD e dos seus parálogos WGD tendem a conservar as proteínas nos mesmos taxa, 41,94% das famílias com MDRs-WGD precedem o aparecimento dos fungos (contra 40,44% das famílias dos seus parálogos WGD). Complementarmente, uma pequena porção similar das famílias com MDRs-WGD (14,61%) e com seus parálogos WGD (13,98%) surgiram no ancestral comum do clado das Saccharomyces (Figura 23).

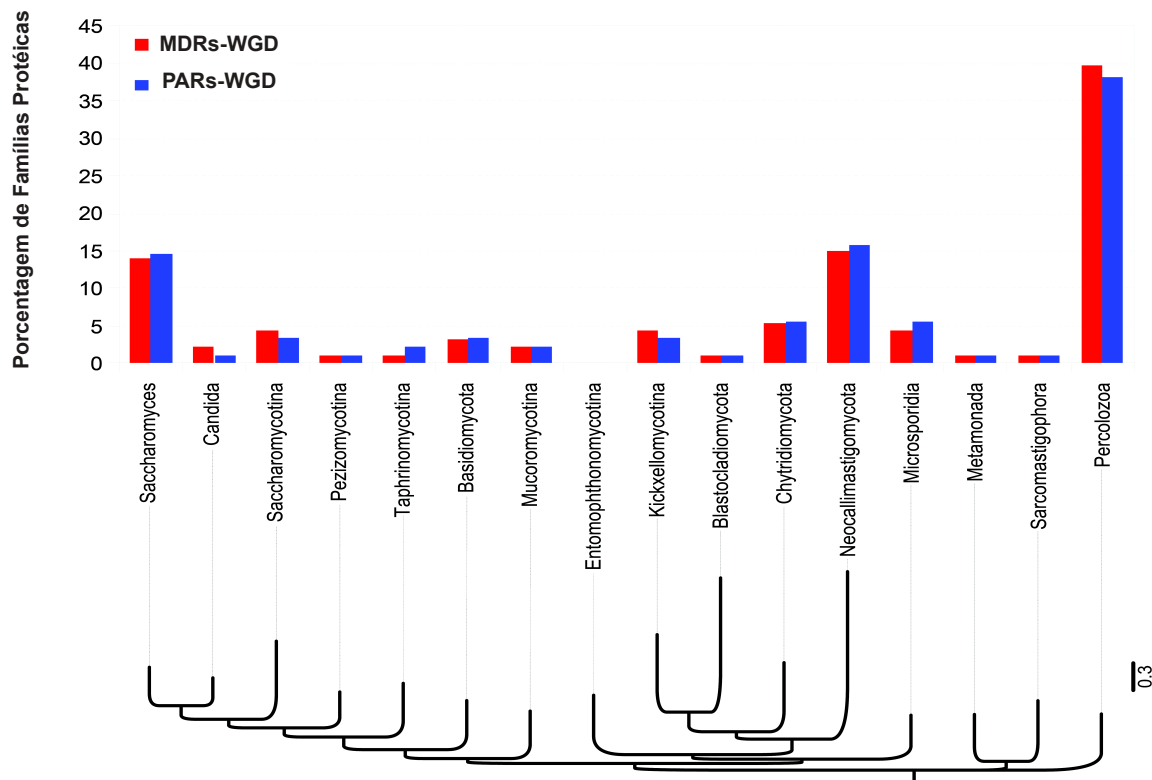


Figura 23: **Classificação taxonômica dos MDRs.** A figura mostra a porcentagem de família protéica por taxon com MDRs-WGD e parálogos WGD. Idem Figura 20.

5.4.5- Os MDRs-SSD são mais conservados que os seus parálogos

As famílias proteicas com MDRs-SSD tendem a serem mais conservadas do que as famílias dos seus parálogos SSD. Especificamente, 66,34% das famílias com MDRs-SSD precedem o surgimento dos fungos (contra 60,76% das famílias com seus parálogos SSD). Reciprocamente, 4,95% das famílias com MDRs-SSD surgiram no clado das Saccharomyces (versus 10,13% das famílias com seus parálogos SSD) (Figura 24).

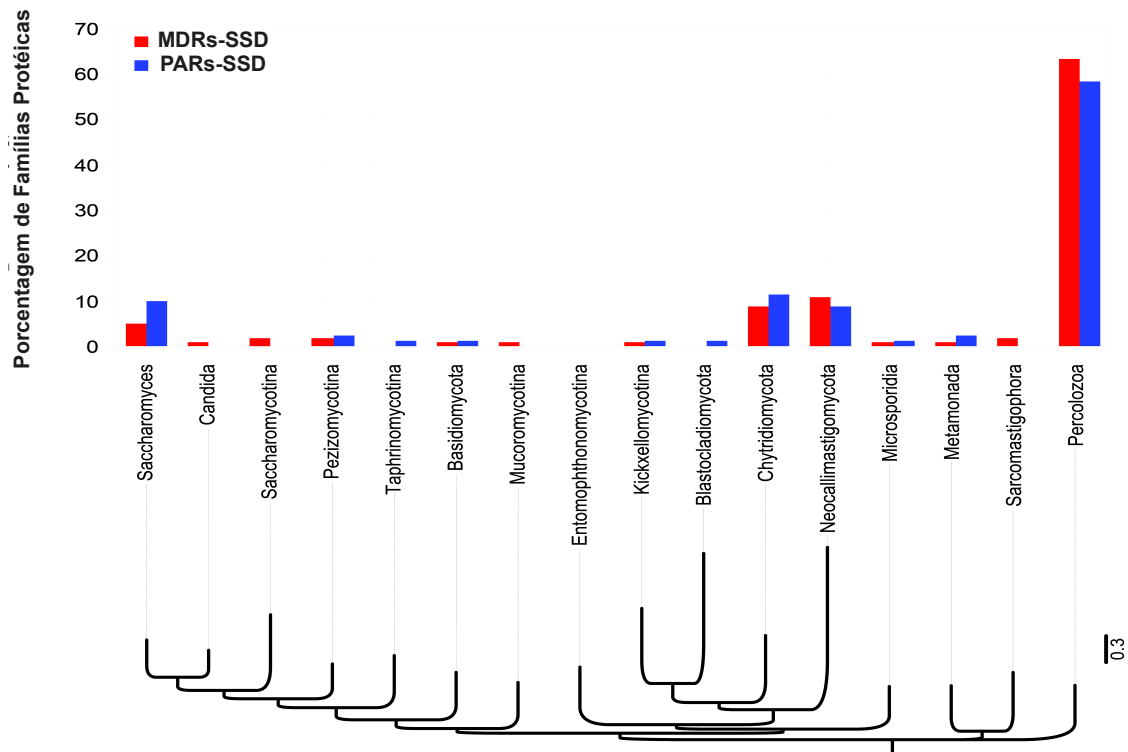


Figura 24: *Classificação taxonômica dos MDRs.* A figura mostra a porcentagem de famílias protéicas por taxon com MDRs-SSD e parálogos SSD. Idem Figura 20.

5.4.6- As famílias protéicas com MDRs-SSD são muito mais conservadas do que as famílias protéicas com MDRs-WGD

A distribuição das famílias com MDRs-SSD e MDRs-WGD através dos taxa é diferente. Especificamente, 66,33% das famílias com MDRs-SSD precedem o surgimento dos fungos (versus 41,92% das famílias com MDRs-WGD). Reciprocamente, 4,95% das famílias com MDRs-SSD surgiram no ancestral comum do clado das Saccharomyces (contra 13,98% das famílias com MDRs-WGD) (Figura 25).

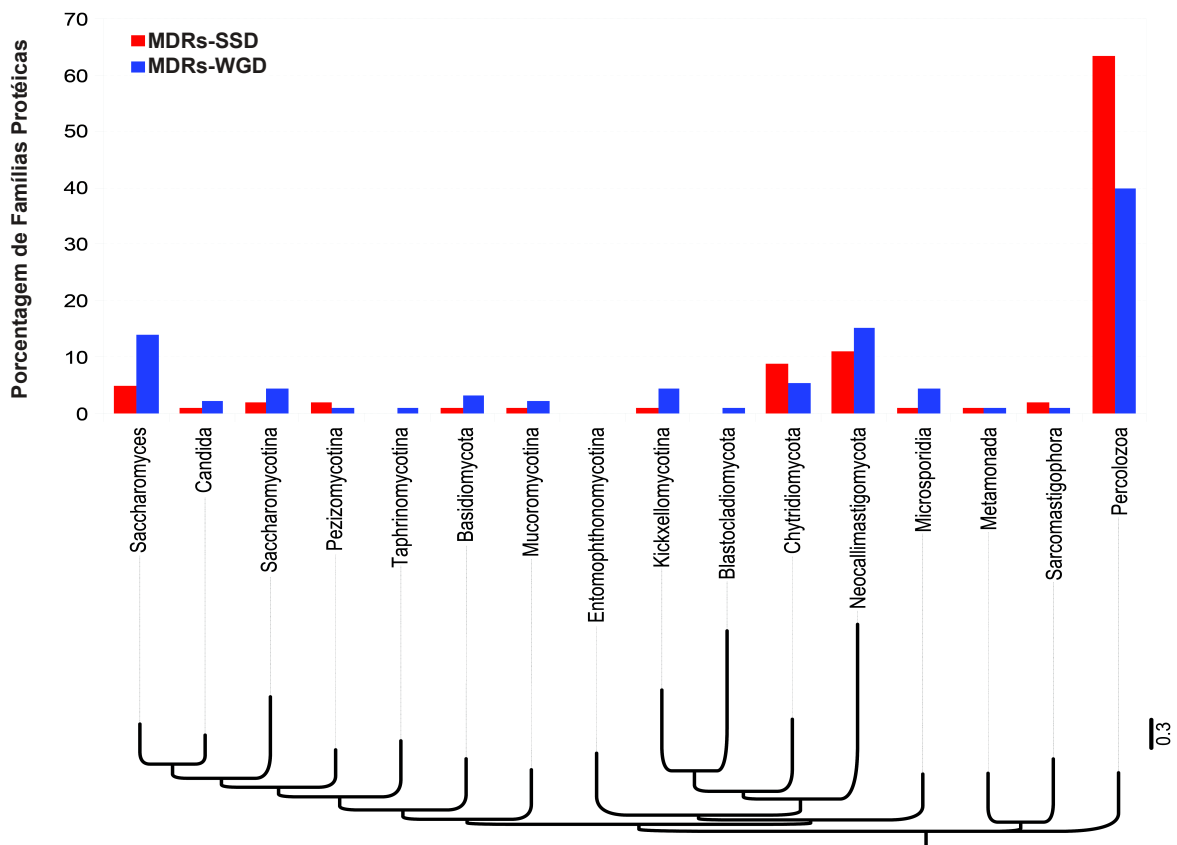


Figura 25: **Classificação taxonômica dos MDRs.** A figura mostra a porcentagem de famílias protéicas por taxon com MDRs-WGD e MDRs-SSD. Idem Figura 20.

6- Discussão

Os dados do agrupamento de proteína (Figura 6) realizado por este presente trabalho corrobora com os dados de Fares *et al.*, 2013. A pequena diferença observada nos resultados das duas metodologias ocorreu porque Fares *et al.*, 2013 utilizou todas as proteínas preditas do genoma da *S. cerevisiae* Sc288 enquanto que este trabalho utilizou apenas proteínas da mesma levedura do RefSeq. O banco de dados RefSeq possui sequencias de referência não redundantes, enquanto que o GenBank pode ser muito redundante para determinados genes. No entanto a metodologia utilizada por Fares *et al.*, 2013 permite identificar genes únicos e parálogos WGD e SSD, enquanto que a metodologia do presente trabalho identifica apenas genes únicos e duplicados.

Ao analisar os genes mutados de *S. cerevisiae* do trabalho de Lang *et al.*, 2013 foi possível identificar que os MDRs são proporcionalmente mais representados entre os “putative drivers” do que os não MDRs (Tabela). Este resultado sugere que os MDRs apresentam grande potencial adaptativo. Além disso, os MDRs sofrem mais mutações ao longo das gerações, possuem maior frequência de mutação (Figura 8) e fixam mais destas mutações do que os não MDRs. Segundo Lang *et al.* 2013 genes com alta taxa de mutação e fixação conferem vantagens adaptativas a população de levedura. Esses resultados sugerem também que apesar dos MDRs possuírem menor quantidade de parálogos que os não MDRs (Figura 6) eles possuem maior robustez mutacional. Portanto os MDRs quando mutados tendem a não prejudicar o fitness da população de *S. cerevisiae*. Os resultados das análises de GIs e PPIs dos MDRs demonstram que apesar dos MDRs sofrerem mais mutações que os não MDRs elas não são deletérias. Observe-se que os MDRs possuem maior número de GIs e PPIs que os não MDRs (Figura 14), portanto maior diversidade funcional. Logo, os MDRs são mais propensos a adquirir novas funções do que os não MDRs. O agrupamento dos MDRs e não MDRs em famílias protéicas indica que os MDRs sofrem mais mutações e possui maior diversidade funcional que os não MDRs, mas eles tendem a ser mais conservados entre as linhagens de fungos que os não MDRs (Figura 20).

O presente trabalho também realizou análises com os subgrupos dMDRs e sMDRs para elucidar o papel da duplicação no destino funcional dos MDRs. Ao

comparar os dMDRs e os sMDRs observou-se que ambos os grupos possuem perfil mutacional e número de GIs e PPIs similares (Figura 15), sugerindo que o evento de duplicação não interfere no destino funcional dos dMDRs. Entretanto o evento de duplicação possui importante papel na conservação dos MDRs através das linhagens fúngicas. Os resultados de agrupamento dos dMDRs e sMDRs em famílias protéicas revelam que os dMDRs são muito mais conservados entre as linhagens fúngica que os sMDRs (Figura 21).

Com objetivo de entender o que ocorre com os dMDRs após a duplicação os domínios dos dMDRs e seus parálogos foram analisados. A maior parte dos dMDRs possuem domínios conservados idênticos (Figura 7) aos seus parálogos, provavelmente conservando-se a função bioquímica. No entanto, há alguma diferença funcional entre as cópias, porque do contrário as associações quimiogénéticas não teriam sido detectadas em virtude da redundância gênica. Os MDRs com parálogos não MDRs possuem proporcionalmente mais pares de proteínas com arquitetura de domínios diferente que os MDRs com parálogos MDRs, sugerindo que os parálogos não MDRs estão perdendo função ou adquirindo novas funções após a duplicação. Além disso também foi observado que os parálogos dos MDRs sofrem mais mutações do que seus respectivos MDRs, permitindo deduzir que o primeiro grupo possui maior robustez a mutação devido à redundância gênica. Entretanto a frequência de mutação é maior nos MDRs do que nos seus parálogos (Figura 10) sugerindo que MDRs quando sofrem mutações conferem mais vantagens adaptativas à população de levedura do que seus parálogos. Os dMDRs também possuem maior número de GIs e PPIs que seus parálogos (Figura 16), conseqüentemente maior diversidade funcional. Além disso os MDRs tendem a ser mais conservados que seus parálogos entre as linhagens fúngicas (Figura 22).

Esse trabalho também elucidou o papel do tipo de evento de duplicação (WGD e SSD) no destino funcional dos MDRs-WGD e MDRs-SSD e seus parálogos. O perfil mutacional através das gerações é similar entre MDRs-WGD e seus parálogos WGD (Figura 11). Entretanto, foi observado que o gene mais mutado no trabalho de Lang *et al.*, 2013 é um parálogo WGD de um MDR, o qual recebeu várias mutações deletérias ao longo das gerações. A observação pontual de que um parálogo WGD de um MDR está recebendo mais mutações e perdendo função após a duplicação corrobora com os resultados das análises de GIs e PPIs. Nessas

análises observou-se que os parálogos WGD dos MDRs-WGD estão perdendo função após a duplicação, porque possuem menor número de GIs e PPIs que os MDRs-WGD (Figura 17). Apesar da provável perda de função após a duplicação os MDRs-WGD e os seus parálogos WGD são conservados de forma similar através dos taxa de fungo (Figura 23).

A duplicação por pequena escala age de forma diferente sobre os MDRs-SSD e seus parálogos. Os MDRs-SSD possuem maior frequência de mutação (Figura 12) e taxa de fixação que os seus parálogos SSD, permitindo deduzir que os MDRs-SSD são os principais responsáveis pela aquisição de novas funções para resistência a drogas. Mas os MDRs-SSD e parálogos SSD não diferem em relação ao número de GIs (Figura 18), sugerindo que após a duplicação os MDRs-SSD estão adquirindo novas funções. Entretanto os MDRs-SSD são mais conservados entre os taxa de fungo que seus parálogos SSD (Figura 24).

Existem 15 pares de MDRs-SSD em que os parálogos também são MDRs, enquanto que existe somente dois pares de MDRs-WGD com parálogos também MDRs. Esses resultados também sugerem que os MDRs após a duplicação SSD estejam adquirindo novas funções por que o número de GIs e PPIs também é semelhante entre os pares. No entanto os parálogos WGD dos MDRs provavelmente estão perdendo função após a duplicação por que além de possuírem poucos pares MDRs eles também possuem menor número de GIs e PPIs que seus MDRs-WGD. Segundo Fares *et al.*, 2013 os parálogos SSD possuem mais GIs que os parálogos WGD. Entretanto, observou-se que as medianas de GIs e PPIs entre MDRs-WGD e MDRs-SSD são semelhantes (Figura 19), sugerindo que a origem dos MDRs duplicados (WGD e SSD) não interfere no seu destino funcional. Entretanto os MDRs-SSD são muito mais conservados entre as linhagens de fungo que os MDRs-WGD (Figura 25).

7- Conclusão

- Muitos MDRs de *S. cerevisiae* surgiram por eventos de duplicação.
- MDRs versus não MDRs:
 - Os MDRs são proporcionalmente mais representados entre os “putative drivers” do que os não MDRs;
 - Os MDRs sofrem mais mutações que os não MDRs;
 - Os MDRs possuem maior diversidade funcional do que os não MDRs;
 - Os MDRs são mais ancestrais do que os não MDRs.
- dMDRs versus sMDRs:
 - A frequência de mutação dos dMDRs é maior do que os sMDRs.
 - Os dMDRs e sMDRs possuem número de GIs e PPIs similar;
 - Os dMDRs são muito mais conservados do que os sMDRs.
- MDRs versus seus parálogos:
 - A maioria dos dMDRs de *S. cerevisiae* possuem domínios protéicos conservados idênticos aos seus parálogos;
 - Os dMDRs sofrem menos mutações que os seus parálogos;
 - Os dMDRs possuem maior diversidade funcional que os seus parálogos;
 - Os dMDRs são mais conservados que seus parálogos.
- MDRs-WGD versus seus parálogos WGD:
 - As mutações através das gerações são similares entre MDRs-WGD e seus parálogos WGD;
 - Os parálogos WGD dos MRDs-WGD estão perdendo função após a duplicação;
 - Os MDRs-WGD e seus parálogos WGD possuem conservação similar através dos taxa.
- MDRs-SSD versus seus parálogos dos SSD:
 - OS MDRs-SSD possuem maior frequência de mutação que seus parálogos SSD;
 - Os MDRs-SSD e seus parálogos SSD estão ganhando novas funções após a duplicação;

- Os MDRs-SSD são mais conservados que os seus parálogos.
- MDRs-WGD versus MDRs-SSD:
 - Os MDRs-WGD e os MDRs-SSD possuem mediana de frequência de mutação similar;
 - Os MDRs-WGD e os MDRs-SSD possuem números similares GIs e PPIs;
 - Os MDRs-SSD são muito mais conservadas do que os MDRs-WGD.

8- Referências Bibliográficas

Alamgir M, Erukova V, Jessulat M, Azizi A, Golshani A: **Chemical-genetic profile analysis of five inhibitory compounds in yeast.** BMC Chem Bio 2010, 10:1-15.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** Nucleic Acids Res 1997, 25:3389-3402.

Barreto L, Canadell D, Petrezselyova S, Navarrete C, Maresova L, Perez-Valle J, Herrera R, Olier I, Giraldo J, Sychrova H, Yenush L, Ramos J, Arino J: **A Genomewide Screen for Tolerance to Cationic Drugs Reveals Genes Important for Potassium Homeostasis in Saccharomyces cerevisiae.** Eukaryotic Cell 2011, 10:1241-1250.

Basehoar AD, Zaton SJ, Pugh BF: **Identification and distinct regulation of yeast TATA box-containing genes.** Cell 2004 116: 699-709.

Blackman RK, Cheung-Ong K, Gebbia M, Proia DA, He S, Kepros J, Jonneaux A, Marchetti P, Kluza J, Rao PE, Wada Y, Giaever G, Nislow C: **Mitochondrial Electron Transport Is the Cellular Target of the Oncology Drug Elesclomol.** PLoS ONE 2012, 7(1):e29798.

Butler G: **Fungal Sex and Pathogenesis.** Clin Microbiol Rev 2010, 23:140-159.

Byrne KP, Wolfe KH: **The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species.** Genome Res 2005, 15:1456-1461.

Castro PA, Savoldi M, Bonatto D, Barros MH, Goldman MHS, Berreta AA, Goldman GH: **Molecular Characterization of Propolis-Induced Cell Death in Saccharomyces cerevisiae.** Eukaryotic Cell 2011, 10:398-411.

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR,

Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED: **Saccharomyces Genome Database: the genomics resource of budding yeast.** Nucleic Acids Res 2012, Jan; 40:700-705.

Conant GC, Wolfe KH: **Turning a hobby into a job: how duplicated genes find new functions.** Nat Rev Genet 2008, 9:938-950.

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, et al. **The genetic landscape of a cell.** Science 2010, 327: 425–431.

Davis-Kaplan SR, Ward DM, Shiflett SL, Kaplan J: **Genome-wide analysis of iron-dependent growth reveals a novel yeast gene required for vacuolar acidification.** J Biol Chem 2004, 279:4322-4329.

de Boer W, Folman WB, Summerbell RC, Boddy L: **Living in a fungal world: The impact of fungi on soil bacterial niche development.** FEMS Microbiol Rev 2005 29:795–811.

Dias PJ, Teixeira MC, Telo JP, Sá-Correia I: **Insights into the mechanisms of toxicity and tolerance to the agricultural fungicide mancozeb in yeast, as suggested by a chemogenomic approach.** OMICS 2010, 14:211-227.

Eddy SR : **A new generation of homology search tools based on probabilistic inference.** Genome Inform 2009, 23:205-211.

Egan MJ, Talbot NJ: **Genomes, free radicals and plant cell invasion: recent developments in plant pathogenic fungi.** Curr. Opin. Plant Biol 2008, 11:367–372.

Engler KH, Coker R, Evans IH: **A novel colorimetric yeast bioassay for detecting trichothecene mycotoxins.** J Microbiol Methods 1999, 35:207-21.

Fares MA, Keane OM, Toft C, Carretero-Paulet L, Jones GW: **The roles of whole-genome and small-scale duplications in the functional specialization of**

Saccharomyces cerevisiae genes. PloS Genet 2013; 9:1-13.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** Genetics 1999, 151:1531-1545.

Gabaldón T, Koonin EV: **Functional and evolutionary implications of gene orthology.** Nat Rev Genet 2013, 14:360-366.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian K, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kötter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang C, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M: **Functional profiling of the Saccharomyces cerevisiae genome.** Nature 2002, 418:387-391.

Gu Z, Nicolae D, Lu HH, Li WH: **Rapid divergence in expression between duplicate genes inferred from microarray data.** Trends Genet 2002, 18:609-613.

Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH: **Role of duplicate genes in genetic robustness against null mutations.** Nature 2003, 421: 63–66.

He X, Zhang J: **Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution.** Genetics 2005, 169:1157-1164.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St

Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G: **The chemical genomic portrait of yeast: uncovering a phenotype for all genes.** Science 2008, 320:362-365.

Hoon S, Gebbia M, Costanzo M, Davis RW, Giaever G, Nislow C: **A Global Perspective of the Genetic Basis for Carbonyl Stress Resistance.** Genes Genom 2011, 1:219-231.

Jaime MD, Lopez-Llorca LV, Conesa A, Lee AY, Proctor M, Heisler LE, Gebbia M, Giaever G, Westwood JT, Nislow C: **Identification of yeast genes that confer resistance to chitosan oligosaccharide (COS) using chemogenomics.** BMC Genom 2012, 13:1-26.

Kapitzky L, Beltrao P, Berens TJ, Gassner N, Zhou C, Wüster A, Wu J, Babu MM, Elledge SJ, Toczyski D, Lokey RS, Krogan NJ: **Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action.** Mol Syst Biol 2010 6:451.

Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*.** Nature 2004 428:617-624.

Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM: **Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations.** Nature 2013, 7464:571-574.

Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM. **Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations.** Nature 2013, 500:571-574.

Learning Perl, 3rd edition. Schwartz RL, Phoenix T; 390 pages; O'Reilly Media, Inc. (2001)

Lee AY, St Onge RP, Proctor MJ, Wallace IM, Nile AH, Spagnuolo PA, Jitkova Y, Gronda M, Wu Y, Kim MK, Cheung-Ong K, Torres NP, Spear ED, Han MK, Schlecht U, Suresh S, Duby G, Heisler LE, Surendra A, Fung E, Urbanus ML, Gebbia M, Lissina E, Miranda M, Chiang JH, Aparicio AM, Zeghouf M, Davis RW, Cherfils J, Boutry M, Kaiser CA, Cummins CL, Trimble WS, Brown GW, Schimmer AD, Bankaitis VA, Nislow C, Bader GD, Giaever G. **Mapping the cellular response to small molecules using chemogenomic fitness signatures.** Science 2014, 344:208-211.

Levy SF, Siegal ML: **Network hubs buffer environmental variation in *Saccharomyces cerevisiae*.** PLoS Biol 2008, 6:264.

Lynch M, O'Hely M, Walsh B, Force A: **The probability of preservation of a newly arisen gene duplicate.** Genetics 2001, 159:1789-1804.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** Nucleic Acids Research 2012, 40:290-301.

Ooi SL, Pan X, Peyser BD, Ye P, Meluh PB, Yuan DS, Irizarry RA, Bader JS, Spencer FA, Boeke JD: **Global synthetic-lethality analysis and yeast functional profiling.** Trends in Genetics 2005, 22:56-63.

Ooi SL, Shoemaker DD, Boeke JD: DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. Nat Genet 2003,35:277-286.

Pan X, Yuan DS, Xiang D, Sookhai-Mahadeo XS, Bader JS, Hieter P, Spencer F, Boeke JD: **A robust toolkit for functional profiling of the yeast genome.** Mol cell 2004, 16:487-496.

Parsons AB, Brost RL, Ding H, Li Z, Zhang C, Sheikh B, Brown GW, Kane PM, Hughes TR, Boone C: **Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways.** Nat Biotechnol 2004, 22:62-69.

Parsons AB, Lopez A, Givoni IE, Williams DE, Gray CA, Porter J, Chua G, Sopko R, Brost RL, Ho C, Wang J, Ketela T, Brenner C, Brill JA, Fernandez GE, Lorenz TC, Payne GS, Ishihara S, Ohya Y, Andrews B, Hughes TR, Frey BJ, Graham TR, Andersen RJ, Boone C: **Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast.** Cell 2006, 126:611-625.

Peyroche G, Saveanu C, Dauplais M, Lazard M, Beuneu F, Decourty L, Malabat C, Jacquier A, Blanquet S, Plateau P: **Sodium Selenide Toxicity Is Mediated by O₂-Dependent DNA Breaks.** PLoS ONE 2012, 7(5):e36343.

Pfaller MA: **Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment.** Am J Med 2012 125:S3-S13.

Pounds JA, Bustamante MR, Coloma L A, Consuegra JA, Fogden MP, Foster PN, La Marca E, Masters KL, Merino-Viteri A, Puschendorf R, Ron SR, Sanchez-Azofeifa GA, Still CJ, Young BE: **Widespread amphibian extinctions from epidemic disease driven by global warming.** Nature 2006, 439:161–167.

Scherens B, Goffeau A: **The uses of genome-wide yeast mutant collections.** Gen Bio 2004, 5:229.

Searls DB: **Pharmacophylogenomics: genes, evolution and drug targets.** Nat Rev Discov 2003, 8:613-623.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets.** Nucleic Acids Res 2006, 34:535-539.

Stijn van Dongen: **Graph Clustering by Flow Simulation.** PhD thesis, University of Utrecht, May 2000. <http://www.library.uu.nl/digiarchief/dip/diss/1895620/inhoud.html>

Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, et al.: **Systematic genetic**

analysis with ordered arrays of yeast deletion mutants. Science 2001, 294: 2364–2368.

Tong AH, Lesage G, Bader GD, Ding H, Xu H, et al.: **Global mapping of the yeast genetic interaction network.** Science 2004, 303: 808-813.

Ulusik I, Kaya A, Unlu ES, Avsar K, Karakaya HC, Yalcin T, Ko A: **Genome-wide identification of genes that play a role in boron stress response in yeast.** Genomics 2011, 97: 106-111.

VanderSluis B, Bellay J, Musso G, Costanzo M, Papp B, Vizeacoumar FJ, Baryshnikova A, Andrews B, Boone C, Myers CL: **Genetic interactions reveal the evolutionary trajectories of duplicate genes.** Mol Syst Biol 2010, 429:1-13.

Venancio TM, Balaji S, Geetha S, Aravind L: **Robustness and evolvability in natural chemical resistance: identification of novel systems properties, biochemical mechanisms and regulatory interactions.** Mol Biosyst 2010, 6:1475-1491.

Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A: **High-resolution yeast phenomics resolves different physiological features in the saline response.** Proc Natl Acad Sci USA 2003, 100:15724-15729.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Véronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW: **Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis.** Science 1999, 285:901-906.

Wiser MJ, Ribeck N, Lenski RE: **Long-Term Dynamics of Adaptation in Asexual Populations**. Science 2013, Epub ahead of print.

Wolfe KH, Shields DC: **Molecular evidence for an ancient duplication of the entire yeast genome**. Nature 1997, 387: 708–713.

Zhang Z, Luo ZW, Kishino H, Kearsley MJ: **Divergence Pattern of Duplicate Genes in Protein-Protein Interactions Follows the Power Law**. Mol Biol Evol 2005, 22:501-505.

9- Apêndice

Tabela S: Proteomas de Referência

Organismos	Abrv	Classificação Taxonômica
<i>Saccharomyces cerevisiae</i>	scer	Saccharomyces_clade
<i>Candida glabrata</i>	cglab	Saccharomyces_clade
<i>Naumovozya castellii</i> (Saccharomscas		Saccharomyces_clade
<i>Eremothecium gossypii</i>	egos	Saccharomyces_clade
<i>Kluyveromyces lactis</i>	klac	Saccharomyces_clade
<i>Candida albicans</i>	calb	Candida_clade
<i>Candida dubliniensis</i>	cdub	Candida_clade
<i>Candida tropicalis</i>	ctro	Candida_clade
<i>Debaryomyces hansenii</i>	dhan	Candida_clade
<i>Meyerozyma guilliermondii</i> (Candida)cgui		Candida_clade
<i>Clavispora lusitanae</i> (Candida lusitana)clus		Candida_clade
<i>Yarrowia lipolytica</i>	yli	Saccharomycotina
<i>Sclerotinia sclerotiorum</i>	sscl	Pezizomycotina
<i>Botryotinia fuckeliana</i>	bfuc	Pezizomycotina
<i>Trichoderma reesei</i>	tree	Pezizomycotina
<i>Chaetomium globosum</i>	cglo	Pezizomycotina
<i>Neurospora crassa</i>	ncra	Pezizomycotina
<i>Phaeosphaeria nodorum</i>	pnod	Pezizomycotina
<i>Aspergillus terreus</i>	ater	Pezizomycotina
<i>Aspergillus oryzae</i>	aory	Pezizomycotina
<i>Aspergillus fumigatus</i>	afum	Pezizomycotina
<i>Emericella nidulans</i> (Aspergillus nidulans)nid		Pezizomycotina
<i>Coccidioides immitis</i>	cimm	Pezizomycotina
<i>Uncinocarpus reesii</i>	uree	Pezizomycotina
<i>Ajellomyces capsulatus</i> (Histoplasma)hcap		Pezizomycotina
<i>Schizosaccharomyces pombe</i>	spom	Taphrinomycotina
<i>Coprinopsis cinerea</i>	ccin	Basidiomycota
<i>Cryptococcus neoformans</i>	cneo	Basidiomycota
<i>Rhizopus oryzae</i>	roro	Mucoromycotina
<i>Mucor circinelloides</i>	mcir	Mucoromycotina
<i>Phycomyces blakesleeanae</i>	pbla	Mucoromycotina
<i>Conidiobolus coronatus</i>	ccor	Entomophthonomycotina
<i>Coemansia reversa</i>	crev	Kickxellomycotina
<i>Catenaria anguillulae</i>	cang	Blastocladiomycota
<i>Batrachomyces dendrobatidis</i>	bden	Chytridiomycota
<i>Gonapodya prolifera</i>	gpro	Chytridiomycota
<i>Piromyces spE2</i>	pspe	Neocallimastigomycota
<i>Edwardsiella aedis</i>	eaed	Microsporidia
<i>Nosema bombycis</i>	nbom	Microsporidia
<i>Encephalitozoon intestinalis</i>	eint	Microsporidia
<i>Trichomonas vaginalis</i>	tvag	Metamonada
<i>Giardia lamblia</i>	glam	Sarcomastigophora
<i>Naegleria gruberi</i>	ngru	Percolozoa