



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Reconstrução in silico da rede metabólica do fungo *Paracoccidioides lutzii*

Paulo Antonio Alvarez

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Orientadora
Prof.^a Maria Emília Walter

Brasília
2012

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Mestrado em Informática

Coordenador: Prof. Ricardo Pezzuol Jacobi

Banca examinadora composta por:

Prof.^a Maria Emília Walter (Orientadora) — Departamento de Ciência da Computação/IE

Prof.^a Maria Sueli Soares Felipe — Instituto de Biologia/UnB

Prof. Ney Lemke — Instituto de Biociências/UNESP

CIP — Catalogação Internacional na Publicação

Alvarez, Paulo Antonio.

Reconstrução in silico da rede metabólica do fungo *Paracoccid-
ioides lutzii* / Paulo Antonio Alvarez. Brasília : UnB, 2012.

78 p. : il. ; 29,5 cm.

Dissertação (Mestrado) — Universidade de Brasília, Brasília,
2012.

1. rede metabólica, 2. *Paracoccidoides lutzii*, 3. bioinformática

CDU 004

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília-DF — Brasil



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Reconstrução in silico da rede metabólica do fungo *Paracoccidioides lutzii*

Paulo Antonio Alvarez

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Prof.^a Maria Emília Walter (Orientadora)
Departamento de Ciência da Computação/IE

Prof.^a Maria Sueli Soares Felipe
Instituto de Biologia/UnB

Prof. Ney Lemke
Instituto de Biociências/UNESP

Prof. Ricardo Pezzuol Jacobi
Coordenador da Mestrado em Informática

Brasília, 22 de fevereiro de 2012

Dedicatória

Dedico este trabalho aos meus pais, a minha família e amigos.

Agradecimentos

Agradeço a minha orientadora pelo apoio durante o desenvolvimento deste projeto. Agradeço a equipe de bioinformática da UnB por todo o suporte prestado. Agradeço a Professora Marie-France Sagot e Cecília Klein pela colaboração durante minha estadia na França.

Sumário

Lista de Figuras	viii
1 Introdução	1
1.1 Contextualização	2
1.2 Problema	4
1.3 Objetivos	4
1.4 Descrição dos capítulos	5
2 Reconstrução de mapas metabólicos	6
2.1 Reações e enzimas	6
2.2 Vias metabólicas	8
2.2.1 Bancos de dados de vias metabólicas	10
2.3 Redes metabólicas	15
2.3.1 Reconstrução de redes metabólicas	16
2.3.2 Ferramentas para criar redes metabólicas	19
2.3.3 Redes metabólicas reconstruídas	22
3 <i>Paracoccidioides lutzii</i>	24
3.1 Descrição geral de fungos	24
3.2 Descrição do <i>Paracoccidioides lutzii</i>	25
3.2.1 O gênero <i>Paracoccidioides</i>	25
3.2.2 O <i>Paracoccidioides lutzii</i>	27
3.2.3 Paracoccidioidomicose	27
3.3 Projeto Genoma Pb	28
4 Métodos	32
4.1 Descrição do método geral	32
4.2 Descrição do método computacional	34
4.2.1 Identificação de enzimas	34
4.2.2 Obtenção de vias metabólicas	36
4.2.3 Construção da rede metabólica	38
4.2.4 Reanotação	40
5 Resultados obtidos e discussão	41
5.1 Resultados	41
5.1.1 Identificação de enzimas e busca de vias nos bancos de dados KEGG e Biocyc	41

5.1.2	Reconstrução da rede metabólica	43
5.1.3	Página com as informações das vias metabólicas do <i>P. lutzii</i> .	43
5.2	Discussão	47
5.2.1	Metabolismo primário	47
5.2.2	Metabolismo secundário	58
5.2.3	Comparação com trabalhos anteriores	59
6	Conclusão e trabalhos futuros	60
	Referências	61

Lista de Figuras

2.1	Exemplo de via metabólica da glicólise, mostrando de forma simplificada as reações envolvidas.	9
2.2	Uma via metabólica circular hipotética.	10
2.3	Exemplo de via metabólica da glicólise, mostrando diferentes compostos de entrada.	11
2.4	Exemplo de via obtida do KEGG mostrando o metabolismo de ácido lipóico [44].	13
2.5	Exemplo de via obtida do Biocyc mostrando a montagem da proteína transportadora de biotina-carboxila [20].	14
2.6	Rede metabólica de <i>Escherichia coli</i> do projeto Ecocyc [47].	15
2.7	Evolução de informações da rede metabólica de <i>E. coli</i> ao longo do tempo [30].	16
2.8	Grafo de compostos das redes metabólicas R_1 , $\{A \rightarrow C, B \rightarrow C\}$, e R_2 , $\{A + B \rightarrow C\}$	19
3.1	Foto de célula de <i>P. brasiliensis</i> com múltiplos brotamentos. Esse formato é conhecido como “roda de leme”. A barra branca representa $10 \mu\text{ m}$. Retirado de [19].	26
3.2	As duas formas do <i>P. lutzii</i> , micélio (a direita) e levedura (a esquerda)	28
3.3	<i>Pipeline</i> de bioinformática do Projeto Genoma Pb.	30
4.1	Método geral de reconstrução da rede metabólica.	33
4.2	<i>Pipeline</i> para a obtenção de vias metabólicas.	35
5.1	Relação entre o número de reações e o número de ORFs na reconstrução da rede metabólica, agrupados por categoria enzimática. . .	42
5.2	Via metabólica encontrada no banco de dados KEGG [44].	43
5.3	Via metabólica encontrada no banco de dados yeastcyc.	44
5.4	Rede metabólica gerada.	45
5.5	Página inicial	46
5.6	Página contendo informações sobre o <i>P. lutzii</i>	46
5.7	Página com dados disponibilizados pelo projeto	47
5.8	Página com <i>links</i> para vias encontradas no KEGG	48
5.9	Página com <i>links</i> para vias encontradas no Biocyc	49
5.10	Reação não encontrada no ciclo TCA. Figura retirada de [65]	51
5.11	Via de síntese da asparagina identificada em <i>P. lutzii</i>	53

Capítulo 1

Introdução

A partir dos estudos feitos em Biologia Molecular [50, 65, 93], verificou-se que as informações necessárias à síntese de proteínas em um dado organismo estão armazenadas no seu DNA, e a passagem do DNA de geração a geração permite que as espécies conservem características essenciais à sua sobrevivência. Dessa forma, a determinação da informação contida no DNA tornou-se um importante meio no estudo das características dos seres vivos.

A informação do DNA pode ser identificada pela sua sequência de bases, sendo feito hoje por meio de técnicas de sequenciamento. No entanto, o trabalho envolvido na determinação de tais bases é, em geral, muito grande. Assim, os projetos genoma, formados por equipes de diversas áreas, visam decodificar e analisar a informação armazenada no DNA. Atualmente, mais de 3.600 projetos genoma foram concluídos e aproximadamente 16.000 estão em andamento [15]. A decodificação das bases de um genoma provê informações que podem ser usadas em diversas áreas do conhecimento como saúde, agricultura e pecuária.

Dentre os diversos projetos genoma já concluídos, podemos destacar o Projeto Genoma Humano (PGH), iniciado em 1992 que visava determinar todos os pares de bases do genoma humano. Após nove anos de trabalho, envolvendo diversas organizações e países, em 2001, foi publicado um rascunho do genoma humano pelo Consórcio Genoma Humano e por uma empresa privada [22, 92]. Além de ter estimulado um grande desenvolvimento das técnicas de sequenciamento de DNA, o PGH impulsionou uma série de outros projetos genoma, com o objetivo de estudar organismos similares ao humano e/ou com características significativas dentro de uma determinada classificação taxonômica.

No Brasil, o primeiro projeto genoma concluído com sucesso foi o sequenciamento do genoma completo da bactéria *Xylella fastidiosa* [83], causadora de uma doença em plantas conhecida como amarelinho, que provoca enormes prejuízos na cultura de laranja, sendo cerca de 30% dos laranjais paulistas afetados por essa doença.

Na Região Centro-Oeste, um grande impulso à área de genômica foi dado com a implantação da Rede Genoma Centro-Oeste. Este projeto foi submetido ao MCT/CNPq, tendo sido aprovado como uma das oito unidades regionais brasileiras para o sequenciamento de genomas. Foi iniciado em 2001 e concluído em 2004, tanto em relação aos experimentos realizados nos laboratórios de biologia molecular quanto

no tocante ao desenvolvimento de um sistema computacional para armazenamento e tratamento das informações biológicas [32, 33]. Após este primeiro projeto genoma, outros foram desenvolvidos, como o Guaraná [6, Projeto Genoma Guaraná] e o Anaplasma [37, Projeto Genoma Anaplasma], além de outro, específico, na área de bioinformática, envolvendo diversas instituições de pesquisa da região Centro-Oeste, denominado Biofoco [Projeto Biofoco].

Até recentemente, os projetos genoma utilizavam principalmente a técnica de sequenciamento Sanger, assim nomeada em homenagem a seu criador Frederick Sanger [79]. Nesta técnica, a molécula de DNA é primeiramente fragmentada em diversos pedaços. Após isso, são feitas diversas cópias dos fragmentos de DNA, seja por meio de reações químicas envolvendo diversas enzimas ou inserindo o fragmento desejado de DNA em vírus ou bactérias e utilizando a capacidade reprodutiva dos mesmos para a realização das diversas cópias necessárias [81]. Após isso, uma técnica conhecida como gel eletroforese é utilizada para a determinação das bases de DNA. Esta última técnica é automatizada, permitindo a determinação das bases de maneira rápida e menos sujeita a erros. Uma vantagem do sequenciamento Sanger é o tamanho dos fragmentos que podem ser sequenciados de cada vez, que pode chegar a até 1000 bases.

Durante duas décadas, esta foi praticamente a única técnica utilizada nestes projetos. Apesar de ser relativamente barata, custando apenas US\$ 0,001 para cada base sequenciada [85], ela torna-se proibitiva quando se tenta sequenciar genomas muito grandes, da ordem de bilhões de bases de DNA. Como em geral o sequenciamento de DNA é feito mais de uma vez, o custo desses projetos se torna muito alto para ser viável com sequenciamento Sanger. Assim, a comunidade científica desenvolveu um novo método mais barato de sequenciamento.

A partir dos anos 2000, novos sequenciadores de alto desempenho foram desenvolvidos para atender a essa necessidade [56]. Estes sequenciadores produzem uma quantidade imensa de dados, a uma fração do custo dos sequenciadores Sanger, por exemplo, laboratórios cobram US\$ 0,0002 para cada base obtida com o sequenciador 454 da Roche. Elas obtém tal desempenho ao realizar o sequenciamento de milhões de sequências de DNA em paralelo, utilizando diversas técnicas, tais como pirosequenciamento e sequenciamento por síntese. Como exemplos desses sequenciadores de alto desempenho podemos citar o 454-FLX da Roche [454], o Illumina [Illumina], e o Solid [SOLiD]. Atualmente, o Brasil dispõe de vários sequenciadores de alto desempenho, como um 454 em São Paulo, outro em Brasília e dois no LNCC (RJ), além de um Illumina localizado em Brasília.

1.1 Contextualização

O *Paracoccidioides brasiliensis* é o fungo causador da paracoccidioidomicose (PCM), uma micose sistêmica com ocorrência em diversos países da América Latina, em especial Brasil, Venezuela e Colômbia, afetando principalmente indivíduos em áreas rurais [5, 86]. O número de pessoas afetadas é estimado em 10 milhões de indivíduos, com cerca de 2% desenvolvendo formas agudas (jovens) ou crônicas (adultos) da doença. A forma aguda da doença compromete principalmente o retículo

endotelial, já a crônica é mais frequente em homens adultos, causando lesões pulmonares ou mucocutâneas. Independentemente do órgão afetado, a PCM frequentemente forma tecido fibroso no pulmão, danificando permanentemente a saúde dos pacientes.

O *P. brasiliensis* ocorre naturalmente na forma de esporo do micélio, e causa a doença após mudar para a forma de levedura no corpo dos pacientes. Este processo de mudança de forma é essencial para o desenvolvimento da doença, pois cepas de *P. brasiliensis* incapazes de efetuar a transição de micélio para levedura não causam a doença. Tal transição é termoregulada e ocorre com a mudança de temperatura experimentada pelo *P. brasiliensis* ao se alojar no corpo de um hospedeiro, sendo tal processo possível de ser observado e revertido em laboratório com a mudança da temperatura de cultivo de 22 para 36° C [32].

A PCM é um problema de saúde pública que afeta trabalhadores de áreas rurais, e não há interesse da indústria farmacêutica em desenvolver fármacos específicos devido ao baixo retorno econômico. A PCM é muitas vezes confundida com tuberculose, sendo os pacientes tratados de forma errônea. Esse fato faz com que frequentemente a PCM só seja identificada em estágio avançado, provocando sérios danos à saúde dos pacientes. Além disso, mesmo quando identificada, a PCM é tratada com antimicóticos de espectro geral, devido à inexistência de drogas específicas.

Dada a alta incidência dessa doença e a existência de especialistas com experiência em pesquisa do *P. brasiliensis* em diversas instituições da Região Centro-Oeste, várias instituições de ensino e pesquisa dessa região criaram em 2001 um projeto para obter informações sobre o transcrito das duas formas do fungo, o projeto Genoma Pb [Projeto Genoma Pb]. Esse projeto tinha como objetivo geral o mapeamento do genoma funcional e diferencial entre as formas de micélio e levedura de *Paracoccidioides brasiliensis*. As coordenações geral e da bioinformática ficaram sob responsabilidade do Instituto de Biologia da UnB, e envolveram diversas instituições de pesquisa desta região, como a Universidade Federal de Goiás, a Universidade Federal de Mato Grosso do Sul e a Universidade Federal de Mato Grosso, entre outras. O projeto Genoma Pb propiciou a formação de recursos humanos em biotecnologia e bioinformática, que impulsionaram a implantação dessas áreas na região Centro-Oeste.

Na etapa de sequenciamento do projeto, feita com a técnica Sanger, foram gerados 19.718 ESTs (*Expressed Sequence Tags*) de alta qualidade. Após a montagem, foram obtidos 2.655 *contigs* e 3.367 *singlets*, constituindo um total 6.022 ESTs montados. O estudo desses dados resultou em um tamanho de genoma estimado em aproximadamente 30 Mb, contendo entre 7.500–9.000 genes [32, 33].

A partir dos dados identificados no Projeto Genoma Pb, vários projetos mais específicos foram desenvolvidos, sobretudo na UnB, na UFG, e em outros grupos no país e no exterior. O Instituto Broad [53], dos Estados Unidos, liderou uma colaboração internacional, particularmente com o Laboratório de Biologia Molecular da UnB, para sequenciar três isolados de *P. brasiliensis*, Pb01, Pb03 e Pb18 [53]. Dados e análises do Projeto Genoma Pb permitiram aprimorar a anotação do DNA de *P. brasiliensis*. Pesquisas posteriores constataram que o isolado do *P. brasiliensis* sequenciado no projeto Genoma Pb era uma espécie distinta dentro do gênero

Paracoccidioides [86], que foi denominado *Paracoccidioides lutzii* [87]. Os demais isolados continuaram sendo chamados de *P. brasiliensis*.

No presente trabalho, utilizaremos os dados do *P. lutzii*, sequenciados pelo Instituto Broad [Broad Institute of Harvard and MIT] dos Estados Unidos. Notamos que o projeto vem sendo desenvolvido em parceria com a Universidade Católica de Brasília (UCB), através da participação direta da Professora Maria Sueli Soares Felipe, da UnB e da UCB, especificamente da área de Biologia Molecular. Além dessa, o projeto iniciou uma parceria com o INRIA/França, particularmente com a Dra Marie-France Sagot e Cecília Klein, ambos da Universidade Claude Bernard (Lyon 1).

Embora milhares de genes do *P. brasiliensis* tenham sido identificados, não existem informações a respeito de como as proteínas codificadas dos genes atuam no sentido de desencadear e regular as reações químicas responsáveis pelo metabolismo do *P. brasiliensis*. Essas informações são muito importantes para o desenvolvimento de drogas para o combate à doença, pois fornecem informação fundamental no sentido de se entender como um dado composto irá afetar o desenvolvimento do fungo. Devido também a falta desses dados e de interesses mercadológicos, hoje não existem drogas específicas para se tratar a PCM, sendo utilizados compostos de amplo espectro para fungos, cujos resultados no tratamento ainda são insatisfatórios.

1.2 Problema

O problema tratado nesta dissertação é que, embora vias metabólicas do *P. lutzii* tenham sido identificadas previamente *in silico* [8], a rede metabólica em larga escala *in silico* do *P. lutzii* ainda não foi reconstruída.

A hipótese do projeto é de que a construção de uma rede metabólica *in silico* para o *P. lutzii* fornecerá informações cujo conteúdo auxiliará no entendimento da doença e no desenvolvimento de drogas para combate à PCM. Mais especificamente, as informações obtidas *in silico* permitirão a identificação de alvos de forma mais rápida e sistemática e auxiliarão na comprovação experimental das informações presentes *in silico*.

1.3 Objetivos

O objetivo principal deste trabalho consiste em reconstruir *in silico* a rede metabólica do fungo *Paracoccidioides lutzii* e disponibilizar informações da rede e de vias metabólicas desse fungo. Disponibilizamos também informações de reações possivelmente encontradas no fungo bem como intermediários metabólicos envolvidos nessas reações.

Temos os seguintes objetivos específicos:

- Identificar vias metabólicas do *P. lutzii* comparadas com *Saccharomyces cerevisiae*, *Coccidioides immitis*, *Escherichia coli*, *Aspergillus fumigatus* e *Aspergillus nidulans*, utilizando o banco de dados KEGG;

- Identificar vias metabólicas do *P. lutzii* comparadas com *S. cerevisiae* e *E. coli*, usando dados do projeto Biocyc;
- Completar, pelo menos parcialmente, as vias identificadas no Biocyc utilizando ferramentas e programas específicos, propondo reanotação para sequências previamente não anotadas, anotadas como proteínas hipotéticas e mesmo com outras anotações;
- Gerar a rede metabólica de *P. lutzii*, identificando as vias encontradas;
- Construir uma página com as informações levantadas e com a rede metabólica do *P. lutzii*;
- Listar intermediários metabólicos e reações encontrados no *P. lutzii*.

1.4 Descrição dos capítulos

No Capítulo 2, discutimos conceitos básicos de metabolismo, necessários ao entendimento deste trabalho.

No Capítulo 3, realizamos uma descrição geral de fungos e caracterizamos de forma mais aprofundada o fungo *P. lutzii*, objeto de estudo deste trabalho.

No Capítulo 4, apresentamos a metodologia geral e o método computacional utilizados para reconstruir a rede metabólica de *P. lutzii*.

No Capítulo 5, inicialmente expomos os resultados obtidos com a metodologia utilizada no capítulo 4 e em seguida discutimos esses resultados e os comparamos com trabalhos anteriores.

No Capítulo 6, concluímos o trabalho e propomos trabalhos futuros.

Capítulo 2

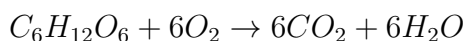
Reconstrução de mapas metabólicos

Neste capítulo, inicialmente apresentamos conceitos básicos de metabolismo necessários ao entendimento deste trabalho. A parte biológica deste capítulo é baseada em [65], obra considerada por estudiosos uma das referências no assunto. Na seção 2.1, apresentamos conceitos relativos a reações químicas e enzimas, fornecendo uma breve explicação de como elas funcionam e outros aspectos importantes de enzimas. Na seção 2.2, definimos vias metabólicas e apresentamos dois exemplos de bancos de dados de vias metabólicas. Na seção 2.3, descrevemos redes metabólicas, fornecemos um exemplo e também abordamos métodos para construção *in silico* de redes metabólicas, como ferramentas computacionais.

2.1 Reações e enzimas

O metabolismo pode ser visto, de forma genérica, como o conjunto de todas as transformações químicas que um ser vivo pode executar. Assim sendo, é necessário compreender alguns aspectos básicos dessas transformações, ou reações químicas.

De forma simplificada, uma reação química consiste na transformação de um conjunto de compostos químicos, denominados reagentes, em um outro conjunto de compostos químicos, denominados produtos. Essa transformação é geralmente apresentada com o uso de uma seta, onde a seta parte dos reagentes e aponta para os produtos. Apresentamos a reação de oxidação da glicose, uma reação muito utilizada pelos seres vivos aeróbicos para obtenção de energia:



Um produto de uma reação pode ser utilizado como reagente de outra reação, isso permite o encadeamento de reações, criando cadeias de processamento. De fato, a reação de oxidação de glicose apresentada anteriormente não é efetuada pelos seres vivos diretamente, mas em várias etapas e envolvendo várias vias metabólicas.

A oxidação da glicose citada anteriormente é uma reação que ocorre muito lentamente. No entanto, essa reação é uma das principais fontes de energia em seres

vivos. Como um ser vivo é capaz de utilizar essa reação para obter energia? Empregando um catalisador, um composto capaz de acelerar uma reação sem ser consumido na reação. Os catalisadores empregados pelos seres vivos são proteínas chamadas enzimas. Algumas enzimas requerem a presença de um ou mais compostos químicos, denominados cofatores, para a catálise da reação, isto é, essas enzimas não são eficientes sem os cofatores. Por exemplo, cofatores podem ser utilizados para transferir íons ou estabilizar determinados compostos durante o processamento da reação.

Enzimas são moléculas altamente específicas. Em geral, uma enzima pode catalisar uma, ou algumas poucas reações, dessa forma um organismo emprega uma grande quantidade de enzimas. Diversas doenças congênitas humanas são causadas pela falta ou inativação de uma enzima, outras são devidas à ação desregulada de enzimas. Várias dessas doenças são fatais, com isso tem-se um argumento a favor da essencialidade do bom funcionamento das enzimas para um organismo poder viver e se reproduzir.

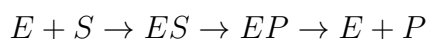
Devido ao grande número de enzimas existentes, bioquímicos adotaram um padrão para nomear as enzimas. Formalmente, uma enzima pode ser identificada por um *Enzyme Commission Number* (EC number ou mais abreviadamente EC). Um EC é um código com quatro campos numéricos associado a uma enzima, onde cada campo fornece informações sobre a reação catalisada por essa enzima. O primeiro campo indica a classe da enzima, dentre as seis possíveis, o segundo número a sub-classe da enzima, o terceiro número informa a subsub-classe da enzima, e o quarto número é um identificador numérico dentro daquela subsub-classe. Normalmente, também é utilizado um nome comum para enzimas, tal como piruvatoquinase, que tem EC 2.7.1.40. O IUBMB [IUBMB] é uma organização internacional que propõe uma classificação para as enzimas em 6 classes [66] (veja tabela 2.1)

Tabela 2.1: Classes de enzimas e função respectiva.

Classe	Função
1	Oxirredutases
2	Transferases
3	Hidrolases
4	Liases
5	Isomerases
6	Ligases

Outra evidência da importância das enzimas é o número de estudos envolvendo essas moléculas. Uma busca pelo termo *enzyme* no *Protein Data Bank* (PDB) [14], por exemplo, retorna milhares de estruturas de enzimas. O estudo da estrutura das enzimas data de 1926.

Uma enzima (E) é capaz de se ligar a um ou mais reagentes de uma reação, denominados substratos (S), catalisar a reação e liberar os produtos (P), esquematicamente:



ES e *EP* são complexos enzima-substrato e enzima-produto. Como se pode ver, as enzimas não são consumidas no curso de uma reação, portanto podem ser utilizadas para catalisar uma nova reação.

Como enzimas aceleram reações químicas, permitem, por exemplo, a oxidação de glicose para obtenção de energia em tempo hábil. No entanto, uma enzima não deve somente ser capaz de catalisar uma reação, ela deve ser capaz de fazê-lo quando necessário, de forma a fazer o melhor uso da energia disponível. Outro aspecto é, se uma enzima capaz de degradar proteínas em aminoácidos agisse todo o tempo, o que impediria essa enzima de digerir proteínas do organismo que não devem ser digeridas? Claramente, a atividade das enzimas deve ser regulada.

2.2 Vias metabólicas

Falamos na seção 2.1 sobre a oxidação da glicose, uma reação muito utilizada para obtenção de energia e feita em diversos passos em organismos aeróbicos. Na verdade, as milhares de reações em uma célula produzem uma infinidade de produtos, e estes podem ter destinos diversos, assim há uma complexa interligação entre as reações de um organismo. Para facilitar o estudo e análise, e também porque é conveniente em alguns casos, certas reações envolvidas em um processo, por exemplo síntese de ácidos graxos, são conceitualmente agrupadas em uma via metabólica. Uma via metabólica em geral consiste num grupo de reações sequenciais capazes de efetuar um processamento específico, em geral síntese ou degradação de um ou mais compostos.

Como uma ferramenta de análise/estudo, uma via metabólica pode ter um nível variado de detalhes: ela pode ter mais ou menos reações, detalhar as reações de um passo ou condensá-las em uma única reação, abordar as reações em um único organismo ou um grupo de organismos, conexão com outras reações/vias, entre muitos outros.

Vias metabólicas são, em geral, mostradas de forma gráfica, e são uma forma de particionar uma rede complexa de interações. Como exemplo, mostramos uma representação da glicólise, uma das primeiras vias estabelecidas e um dos primeiros passos na oxidação da glicose. A glicose tem como produto “final” o piruvato, esta molécula pode ser utilizada para fermentação, oxidada para gás carbônico e água, entre outros propósitos. Como se vê, as reações em um organismo possuem interligações, uma via metabólica é um meio para visualizar/estudar um grupo de interesse.

A figura 2.1 é uma representação simplificada da via da glicólise, que quebra a glicose e obtém duas moléculas de piruvato ao final. Pela figura, a glicólise consiste em um encadeamento linear de reações, por esse mesmo motivo, vias com essa característica são chamadas de lineares. Um outro tipo de via clássica são as chamadas vias circulares, onde ocorre um ciclo de reações, produzindo a cada “volta” um conjunto específico de moléculas. A figura 2.2 mostra um formato de via circular. Como dissemos anteriormente, vias metabólicas são uma representação conveniente, mas não necessariamente real. Para ilustrar esse ponto, apresentamos uma visão diferente da glicólise, na figura 2.3, mostrando não apenas a

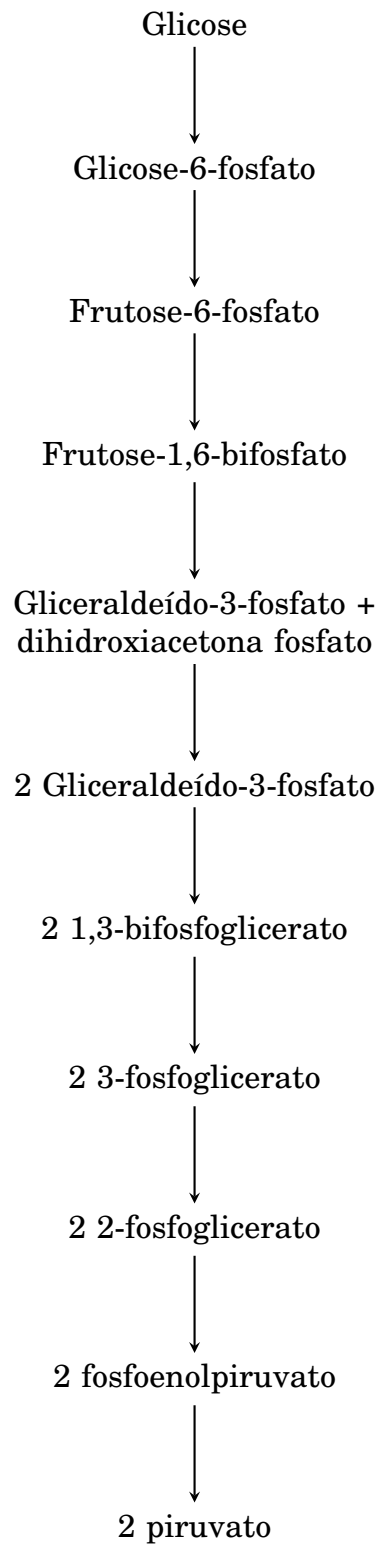


Figura 2.1: Exemplo de via metabólica da glicólise, mostrando de forma simplificada as reações envolvidas.

glicose, mas outros carboidratos e compostos capazes de participar em alguma das reações da via. Como se vê, com esse nível de detalhes, a via fica mais complexa. Considerando que alguns destes compostos podem ser obtidos por outras vias, a glicólise pode ficar bem mais complicada.

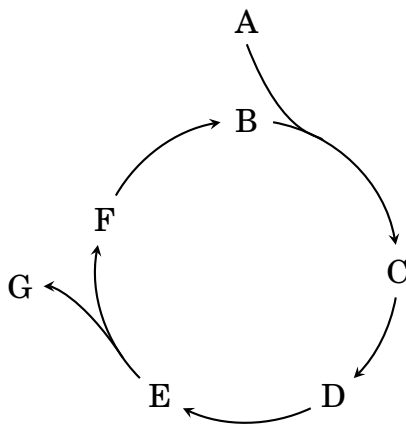


Figura 2.2: Uma via metabólica circular hipotética.

2.2.1 Bancos de dados de vias metabólicas

Ao longo de mais de um século de pesquisa, muita informação sobre metabolismo já foi acumulada, e também várias vias metabólicas já foram descritas e comprovadas experimentalmente. Existem diversos bancos de dados contendo informações não só de vias metabólicas, como de reações, compostos, organismos e drogas, entre outros. Esses bancos permitem a visualização das vias, e em geral oferecem *links* para páginas contendo mais informações sobre um dado aspecto. Os bancos também podem oferecer uma busca, por exemplo, dado um *EC number*, o banco retorna todas as vias onde esse EC ocorre, alguns permitem filtragem por organismo, entre outros.

Alguns desses bancos de dados podem ser acessados via internet, e contêm uma variedade de informações. Alguns têm seu foco em vias mais gerais aplicadas a vários organismos, enquanto outros focam nas vias de um ou poucos organismos. A principal forma de interação com esses bancos consiste na busca de vias contendo *EC numbers*. Alguns bancos permitem buscar uma lista de ECs, retornando as vias com pelo menos um EC da lista, outros fazem buscas apenas por ECs individuais. Alguns bancos retornam figuras de vias com os ECs encontrados destacados graficamente, outros bancos retornam apenas a figura das vias. Existem diversas outras características, cada qual voltada a um aspecto em particular.

Neste trabalho, foram utilizados dois bancos de dados de vias metabólicas públicos, o Kyoto Encyclopedia of Genes and Genomes (KEGG) e os bancos de dados do projeto Biocyc. Em seguida, detalharemos um pouco mais esses bancos.

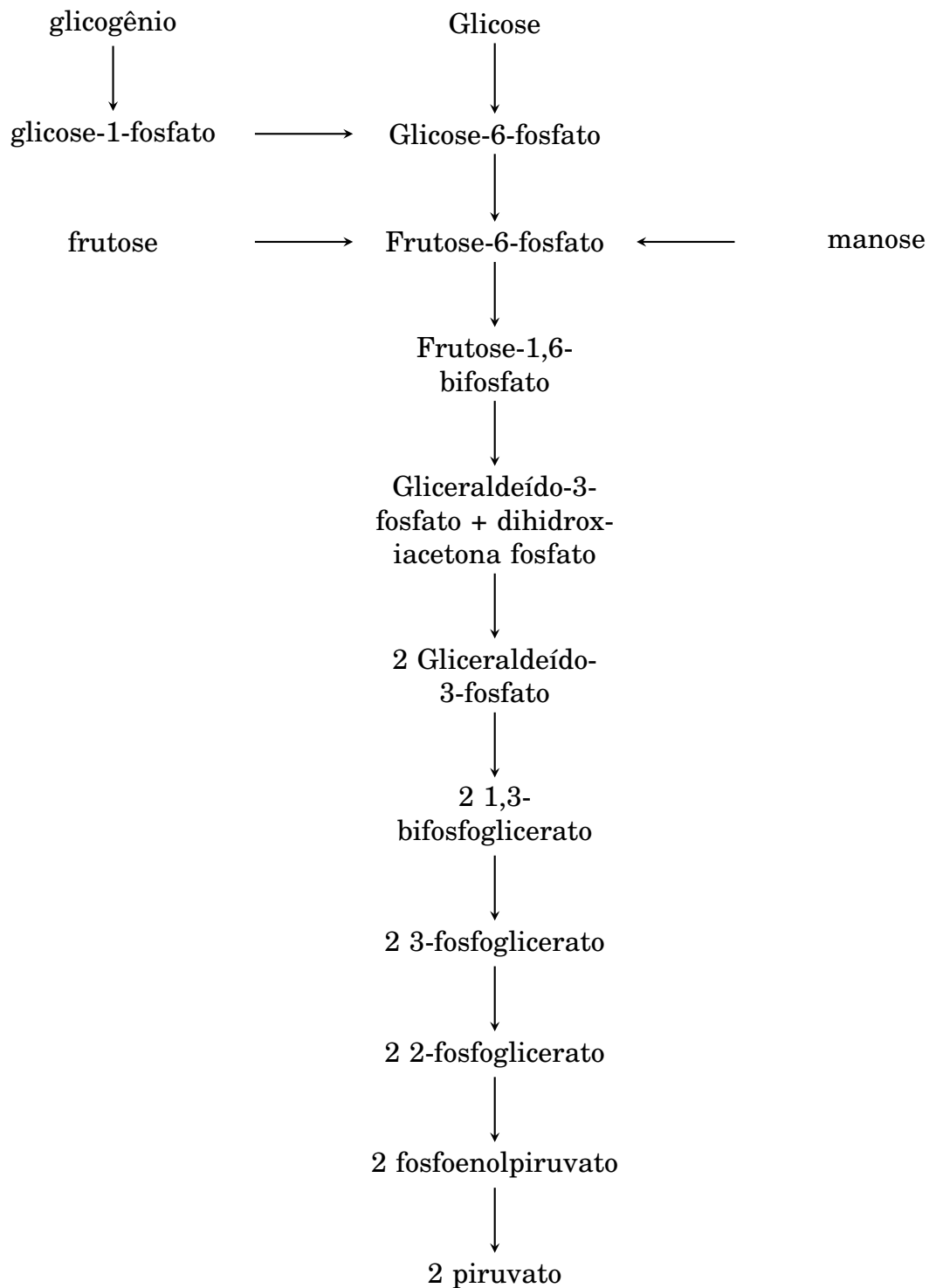


Figura 2.3: Exemplo de via metabólica da glicólise, mostrando diferentes compostos de entrada.

KEGG

O KEGG [44] é um grande banco de dados contendo informações não somente de vias metabólicas, mas também de reações, compostos, genes, genomas, drogas, ortologia entre outros, sendo portanto muito utilizado.

O KEGG permite o envio de um arquivo contendo em cada linha a especificação de um EC, e busca nas vias metabólicas de um organismo as vias onde pelo menos um EC do arquivo ocorre. O resultado contém não apenas uma lista de *links* para as vias encontradas como também os ECs não encontrados na busca. Os ECs encontrados nas vias são destacados colorindo-se uma caixa com uma cor que pode ser especificada pelo usuário. A figura 2.4 mostra uma via metabólica do KEGG, ECs encontrados nessa via são coloridos com uma borda vermelha. ECs com fundo branco não estão presentes ou não foram identificados no organismo cujas vias são pesquisadas.

O KEGG também oferece a ferramenta KAAS (KEGG Automatic Annotation Server) [64], para anotação de genomas utilizando bancos de dados de sequências ortólogas. De forma simplificada, a ferramenta toma como entrada um conjunto de sequências para ser anotado, e alguns outros parâmetros. Essas sequências são então comparadas por BLAST com um conjunto de sequências de referência. Os homólogos identificados e com pontuação acima de um limiar são tomados como candidatos a ortólogos. Os ortólogos são divididos em grupos KEGG Orthology (KO). De forma resumida, o KO é um conjunto de genes ortólogos construídos a partir de vias. Calcula-se para cada ortólogo uma pontuação de atribuição baseada em heurísticas para cada KO. O número identificador do KO com maior pontuação é atribuído para a sequência.

Biocyc

O projeto Biocyc [20] consiste em uma grande coleção de bancos de dados, cada um dedicado a um organismo. Esses bancos são divididos em três níveis, de acordo com a quantidade de revisão manual utilizada em cada banco. No primeiro nível estão presentes os bancos com extensa revisão manual por especialistas, esses bancos em geral contém muitas referências a artigos com comprovações experimentais dos dados do banco e são constantemente revisados por equipes com experiência no organismo em questão. Por demandar uma grande quantidade de trabalho, esses bancos também são em número pequeno, cerca de 6. No segundo nível, estão presentes os bancos de dados de vias metabólicas com alguma revisão manual, esses bancos em geral são bem menos curados comparados aos do primeiro nível. Por fim, no terceiro nível, estão os bancos de dados contendo apenas previsões computacionais, sem nenhuma revisão manual significativa, esses bancos constituem a maior parte dos bancos de dados disponíveis. A figura 2.5 mostra uma via obtida a partir do Biocyc.

O Biocyc utiliza um banco de dados conhecido como MetaCyc [20], que armazena vias metabólicas comprovadas experimentalmente. A figura 2.5 é um exemplo de via metabólica armazenada no MetaCyc.

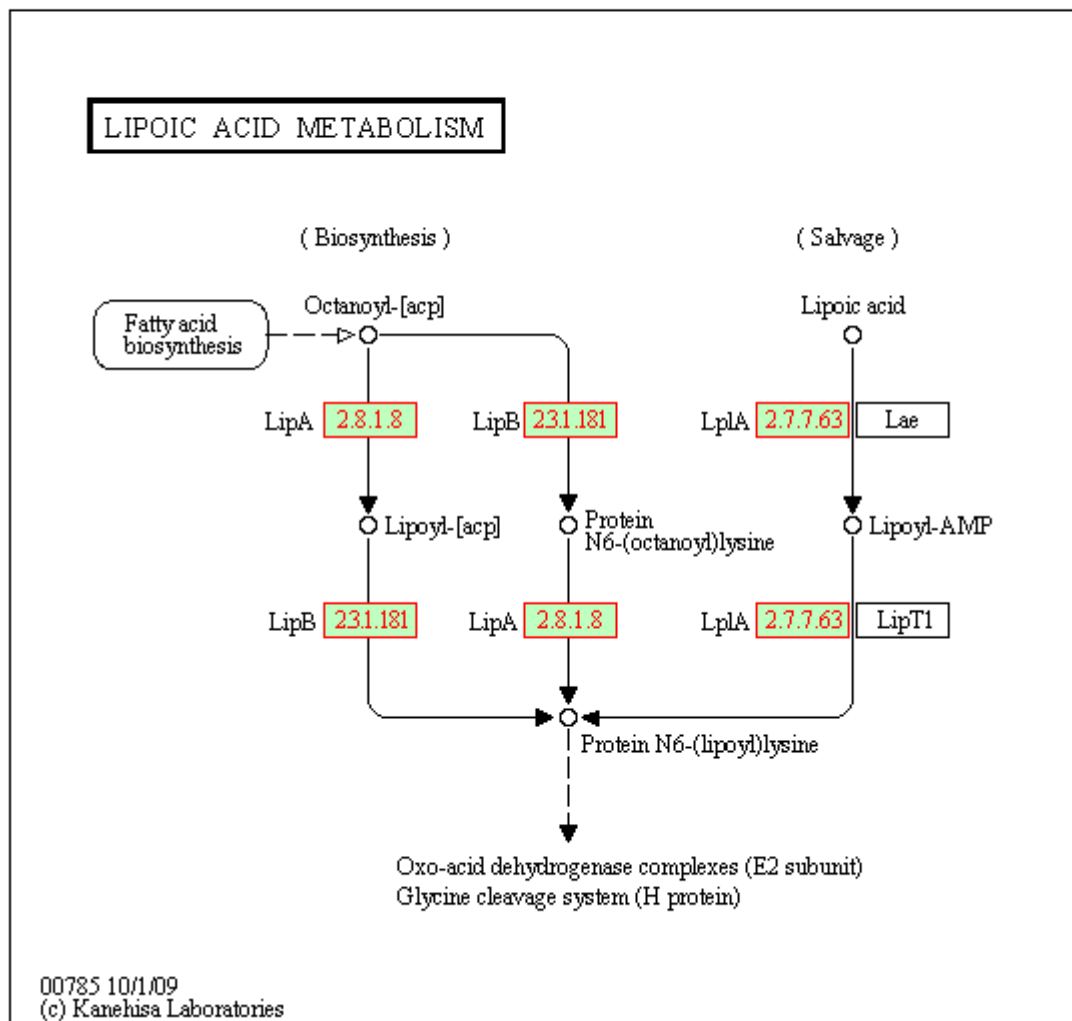


Figura 2.4: Exemplo de via obtida do KEGG mostrando o metabolismo de ácido lipóico [44].

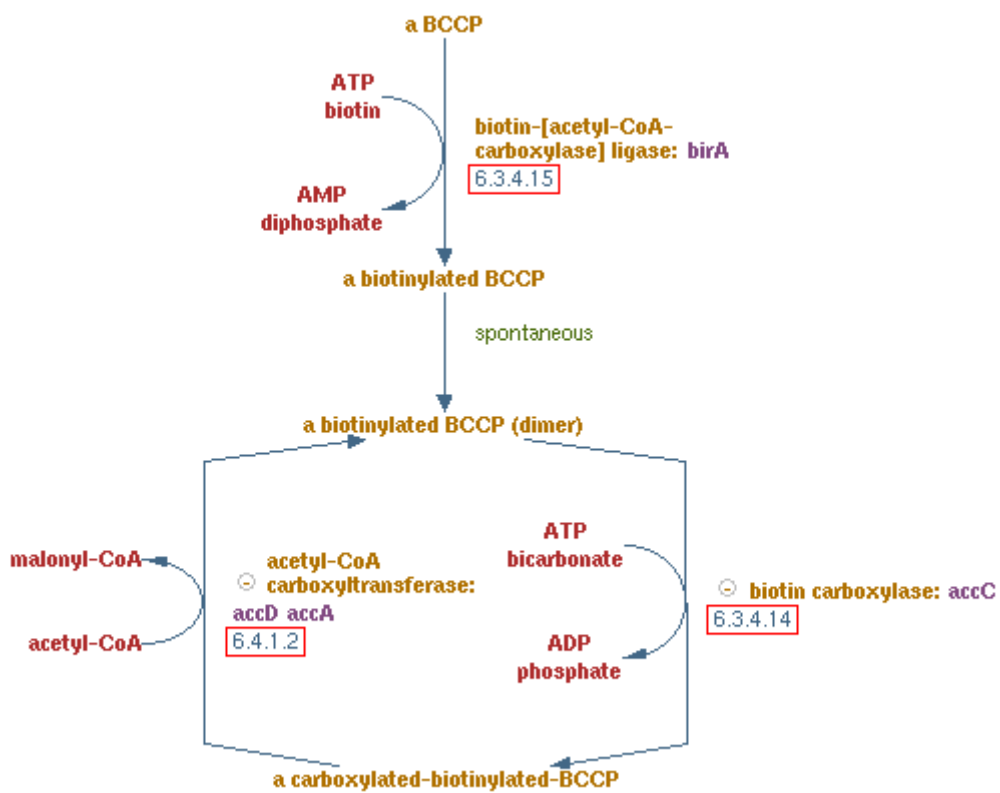


Figura 2.5: Exemplo de via obtida do Biocyc mostrando a montagem da proteína transportadora de biotina-carboxila [20].

2.3 Redes metabólicas

Uma rede metabólica pode ser entendida como o conjunto de todas as reações efetuadas por um organismo. Assim como as vias metabólicas, uma rede metabólica pode apresentar um nível variado de detalhes, como mais ou menos reações, informações de regulação ou não, informação espacial das reações, como compartimento celular, entre outros. A figura 2.6 mostra uma visualização da rede metabólica de *Escherichia coli* (*E. coli*) do projeto Ecocyc [46], os retângulos representam compartimentos celulares, as linhas azuis são reações conectando compostos químicos. Como pode ser visto na parte direita da figura, existem algumas reações não agrupadas em vias, e inclusive algumas reações desconectadas do resto da rede. O nível de detalhes da rede pode ser mudado, permitindo visualizar detalhes não mostrados na figura 2.6.

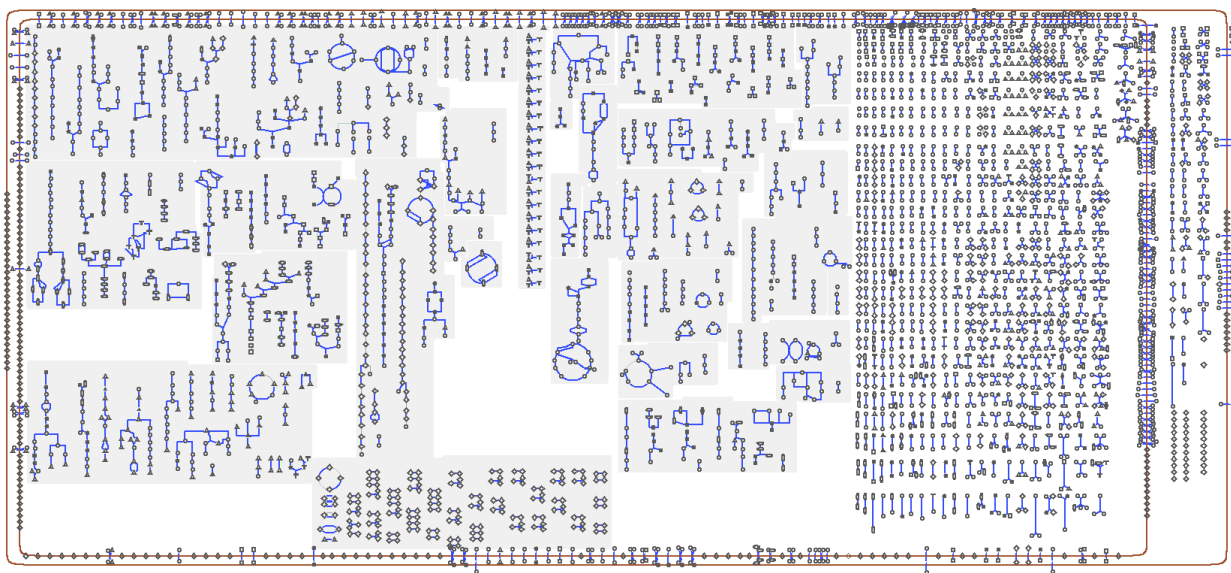


Figura 2.6: Rede metabólica de *Escherichia coli* do projeto Ecocyc [47].

Uma rede metabólica é, em geral, incompleta, por exemplo, reações podem faltar na rede, reações podem não ter um EC associado, ou uma reação pode ter sido detectada no organismo mas o gene responsável pela produção da enzima catalisadora da reação é desconhecido, entre outros. De fato, mesmo a rede metabólica de *E. coli* não é “completa”, com várias evoluções ao longo do tempo. A figura 2.7, retirada de Feist et al [30], mostra a evolução de informações da rede metabólica de *E. coli* ao longo do tempo até o ano de 2007. Como se pode ver, a construção de uma rede metabólica é iterativa, com cada estudo aprofundando e melhorando os dados dos estudos anteriores.

Redes metabólicas podem ser usadas para uma variedade de propósitos: para melhorar a anotação de um genoma ao identificar possíveis erros e inconsistências, estudar a evolução das espécies pois espera-se que organismos próximos tenham metabolismos muito parecidos, o estudo das diferenças ajuda a revelar quais fatores influenciam a capacidade de um organismo, como a capacidade de infectar

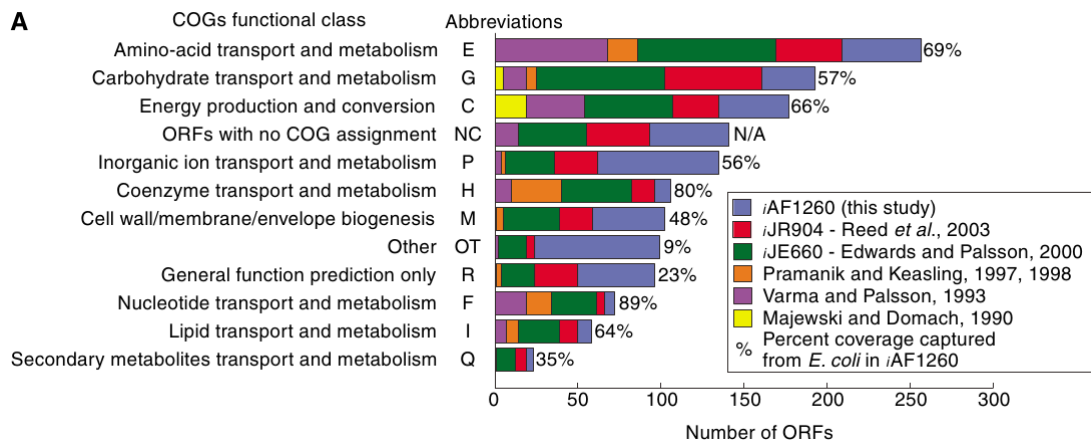


Figura 2.7: Evolução de informações da rede metabólica de *E. coli* ao longo do tempo [30].

um certo hospedeiro. Redes metabólicas também podem ser utilizadas no desenvolvimento de drogas, identificando pontos essenciais do metabolismo e inibindo-os, pode-se matar ou debilitar seriamente um organismo. Podem também ser utilizadas em engenharia genética, fornecendo dados de como o metabolismo pode ser afetado com introdução/remoção de alguma capacidade.

2.3.1 Reconstrução de redes metabólicas

A reconstrução de uma rede metabólica é um processo laborioso, demorado e com diversos detalhes a serem considerados. A princípio, o mínimo necessário para uma rede metabólica consiste no conjunto de reações de um organismo. Esse conjunto pode ser obtido a partir de uma extensa revisão da literatura disponível, não somente de artigos, mas também de livros, bancos de dados já existentes, entre outras fontes, um exemplo de trabalho empregando esse tipo de abordagem pode ser visto em Förster et al [35]. Esse tipo de trabalho é mais diretamente aplicável em organismos modelo, com biologia bem conhecido e ampla evidência experimental disponível [31]. Como exemplo citamos a levedura *S. cerevisiae*, existe um banco de 2.026 mutantes, cada um dos mutantes tem um gene inativo. Por esse estudo, descobriram-se vários genes essenciais [94].

Esse tipo de reconstrução é um processo com alto trabalho manual, mas caso dados experimentais sejam a principal evidência das reações incluídas, a rede metabólica gerada pode ter uma alta qualidade, no sentido da existência de grande parte das reações ser justificada por estudos com resultados experimentais.

Por causa do grande trabalho envolvido, das diversas áreas do conhecimento abrangidas, alguns projetos adotaram, entre outros fatores, um processo baseado em comunidades para construção de redes metabólicas. Nesse formato, uma comunidade de especialistas de várias áreas se reúne para discutir, planejar, executar e curar a reconstrução de uma rede metabólica. Esse processo tem o potencial de agregar o conhecimento de muitas pessoas, e assim permitir a criação de uma rede

metabólica de alta qualidade, mas apresenta problemas, dentre eles a necessidade de disponibilizar os dados para a comunidade, se possível em um formato padrão. A criação do padrão *Systems Biology Markup Language* (SBML) [41] em 2003 representa um avanço nessa direção, pois muitas redes metabólicas são disponibilizadas nesse formato e uma série de ferramentas capaz de processá-lo existe. Um exemplo de iniciativa envolvendo a construção de uma rede metabólica com o auxílio de uma comunidade pode ser visto em Hergard et al [39]. Como evidência da constante evolução de informações em uma rede metabólica, outros estudos foram publicados com a atualização dessa rede, como Dobson et al [26] e Heavner et al [38].

Para muitos organismos, no entanto, a quantidade de dados bioquímicos disponíveis na literatura é baixa ou talvez até mesmo inexistente, se considerarmos um organismo recém-descoberto por exemplo. Embora o ideal seja a criação de uma rede metabólica baseada em dados experimentais, os experimentos bioquímicos, genéticos, de sequenciamento e outros necessários para obter evidência experimental são caros, demorados e/ou de execução complicada, assim, para a maioria dos organismos simplesmente não é viável a construção de uma rede metabólica baseada na maior parte em sólidos resultados experimentais.

No entanto, é cada vez mais fácil sequenciar o genoma de um organismo, e pela aplicação de uma série de métodos computacionais obter um conjunto de transcritos candidatos desse organismo. Esses transcritos podem ser posteriormente analisados por métodos como Blast [3], PFAM [13] para obter os ECs possivelmente associados com cada transcrito, ou ainda outros métodos comparativos. O resultado dessa atribuição de ECs a transcritos seria então a base para construção da rede metabólica. Embora esse processo possa ser executado de forma relativamente rápida, vários erros podem ser introduzidos: um possível transcrito pode nunca ser traduzido em proteína, portanto a reação catalisada pela enzima correspondente não existe na rede metabólica real, algumas das reações encontradas de forma comparativa podem não ocorrer realmente no organismo, consistindo erros do método comparativo. Mesmo quando uma reação realmente existe em um organismo, as informações sobre uma enzima e a correspondente reação catalisada podem não ser completas, como exemplo, a enzima álcool desidrogenase pode aceitar mais de um tipo de álcool como substrato, e em alguns casos a reação catalisada é descrita como: 1 álcool + NAD resultando em 1 aldeído + NADH. Por causa da possibilidade de erros, é desejável a revisão por especialistas das reações obtidas.

Feita essa etapa de coleta das reações catalisadas em um organismo, esse conhecimento deve ser transformado em um modelo matemático para permitir análises e validações posteriores [31]. Uma possível análise consiste no chamado *flux balance analysis* (FBA). Nessa análise, a rede metabólica é modelada por meio de uma matriz estequiométrica S com m linhas e n colunas, onde cada linha representa um composto dos m existentes e cada coluna uma das n reações. Em cada coluna, os números representam os coeficientes estequiométricos dos compostos participantes da reação, valores negativos representam compostos consumidos (reagentes), valores positivos compostos produzidos (produtos), 0 representa um composto que não participa da reação. O fluxo de cada reação na rede é representado por um vetor v de tamanho n , para cada reação i , $v[i]$ representa a quantidade de matéria fluindo por aquela reação. Assumindo um estado de equilíbrio, onde todo material

produzido é consumido, obtemos a seguinte equação:

$$Sv = 0$$

Se $v = 0$, a equação é trivialmente satisfeita, no entanto, se pelo menos um valor de v for maior que zero, teremos um sistema de equações com m equações e n incógnitas. Em geral, temos mais reações em um organismo que compostos, então $n > m$, com mais incógnitas que equações, o sistema $Sv = 0$ tem mais de uma solução. É possível restringir o espaço de soluções restringindo os valores de v para representar o fluxo máximo e mínimo de uma reação observada em um organismo, por exemplo, podemos definir $1 \leq v[3] \leq 5$. Também podemos definir c , um vetor de pesos, onde cada $c[j]$ indica quanto a reação j contribui para uma função objetivo, dessa forma, temos $Z = c^T v$ uma função objetivo. Usando programação linear, pode-se procurar um ponto no espaço de soluções restrito de $Sv = 0$ que maximiza/minimiza a função objetivo Z . Um exemplo de aplicação desse método consistiria em definir uma função objetivo representando a geração de biomassa, ou crescimento, do organismo. Comparando-se a solução obtida por FBA com um experimento de crescimento do organismo em um dado meio é possível detectar inconsistências no modelo. Um valor máximo de produção de biomassa de 0 indicaria que o organismo não é capaz de crescer em um dado meio, se isso for contrário à evidência experimental, indica que a rede metabólica está incompleta, faltam reações capazes de permitir o crescimento em um dado meio. Se, pelo contrário, a análise prevê crescimento em um meio onde o organismo não pode crescer, isso é um indicativo que a rede contém reações que não estão presentes no organismo, e devem ser removidas.

Um outro modelo de redes metabólicas consiste no uso de grafos. Um grafo G é constituído de dois conjuntos: V e E . O conjunto V é o conjunto de vértices, e $E \subseteq V^2$ é o conjunto de arestas, onde um elemento $(a, b) \in E$ representa uma aresta saindo do vértice a e indo em direção ao vértice b . Existem diversas formas de se modelar o conhecimento das reações de um organismo em um grafo. No grafo de compostos, cada vértice representa um composto, e existe um vértice do composto a para o composto b se existe uma reação onde a é consumido e b é gerado. No grafo de reações, cada vértice representa uma reação, e existe uma aresta da reação R_x para a reação R_y se pelo menos um dos produtos de R_x for um dos reagentes de R_y . No grafo bipartido de reações e compostos, o conjunto de vértices V é dividido em dois subconjuntos, C o subconjunto de compostos e R o subconjunto de reações, tais que não existe intersecção entre eles, isto é, $C \cap R = \emptyset$. Sejam $c_i \in C$ e $r_j \in R$, uma aresta no grafo admite apenas duas formas, ou (c_i, r_j) , representando que o composto c_i é consumido na reação r_j , ou (r_j, c_i) , representando que a reação r_j produz o composto c_i . Todos esses modelos podem gerar a mesma representação em grafo para duas redes metabólicas distintas, gerando assim uma ambiguidade. Um exemplo é mostrado na figura 2.8.

A análise de uma rede metabólica representada no formato de um grafo permite a detecção de compostos bloqueados, definidos como compostos nunca produzidos ou nunca consumidos. Existem programas para detectar esses compostos e também para tentar corrigir essas falhas, buscando, por exemplo, o menor número de

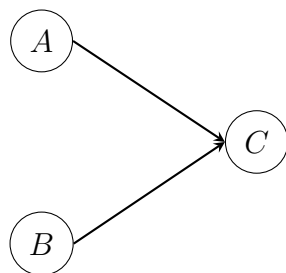


Figura 2.8: Grafo de compostos das redes metabólicas R_1 , $\{A \rightarrow C, B \rightarrow C\}$, e R_2 , $\{A + B \rightarrow C\}$.

operações de adição de reação, reversão de direções de reações e transformação de uma reação unidirecional em uma reação bidirecional capaz de eliminar esses compostos bloqueados. Também é possível realizar análises de predição de alvos de drogas, levando em conta inclusive a biologia de um hospedeiro, ao eliminar por exemplo reações vitais para um hospedeiro da lista de reações previstas para alvos de drogas [24].

Uma revisão do processo de reconstrução de redes metabólicas é feita em [31]. Inicialmente, as informações de enzimas eram obtidas por caracterização bioquímica e revisão de literatura, esse método manual era aplicado para alguns organismos seletos. Com novas ferramentas capazes de anotar genomas inteiros rapidamente, esse processo pode ser aplicado para outros organismos. O processo de reconstrução de uma rede metabólica possui quatro grandes passos: primeiramente realiza-se a anotação do genoma e identifica-se as enzimas e reações catalisadas pelas enzimas. Em seguida essas informações são revistas para identificar erros e omissões, com a ajuda da literatura e especialistas no organismo em questão. Com as informações revistas, elas são representadas em um formato matemático, permitindo novas análises, inclusive de consistência. Por último, a reconstrução pode ser usada, e dados de sequenciamento de alto desempenho ajudam a melhorar a rede.

Para este trabalho, utilizamos como base uma reconstrução da rede metabólica de *S. cerevisiae* [35]. Essa foi a primeira reconstrução de uma rede metabólica de uma célula eucariótica, além disso, *S. cerevisiae* é um fungo, oferecendo assim uma maior proximidade taxonômica em relação ao *P. lutzii*. Essa rede foi reconstruída utilizando principalmente informações da literatura, extensas para a *S. cerevisiae*. Embora a literatura disponível sobre *P. lutzii* não seja tão abundante, o método exposto no artigo, de integração de dados de diferentes fontes para a construção da rede metabólica, serviu de base para este trabalho.

2.3.2 Ferramentas para criar redes metabólicas

Existem ferramentas para auxiliar no processo de criação e análise de uma rede metabólica, algumas ferramentas são dedicadas somente à análise, outras somente à criação, e outras ainda incorporam ambas funções. Apresentamos uma breve descrição de algumas das ferramentas disponíveis.

O método PRIAM [21] trabalha obtendo dados do banco de dados ENZYME [12]. Para cada EC disponível no ENZYME, as sequências proteicas correspondentes a cada EC são retiradas e utilizadas para a criação de um perfil de cada um dos ECs. Para cada grupo de sequências, módulos capazes de caracterizar o grupo são detectados, e após isso é feita uma seleção dos módulos essenciais. Por fim, para levar em conta a possibilidade de enzimas não-homólogas catalisando uma mesma reação e outros casos, esses módulos essenciais são combinados utilizando-se regras lógicas do tipo E e OU. Por exemplo, um dado EC pode ter quatro módulos detectados, M_1 , M_2 , M_3 e M_4 , para uma sequência ser classificada como tendo esse EC, ela deve ter os módulos M_1 e M_2 ou ter os módulos M_3 e M_4 , sendo assim a regra correspondente seria: $(M_1 \text{ E } M_2) \text{ OU } (M_3 \text{ E } M_4)$. A execução consiste em comparar cada proteína com os perfis gerados, reportar os melhores resultados não-sobrepostos para cada proteína, obtendo por fim uma lista de ECs preditos. Este programa realiza somente a predição de reações, não realizando nenhuma outra forma de análise posterior.

A ferramenta metaSHARK [69] possui o módulo SHARKHunt inicia com a preparação de dois conjuntos de perfis: um de polipeptídeos PSI-BLAST [4] e outro de cadeias escondidas de Markov do programa HMMER. Foram utilizados o conjunto de perfis PSI-BLAST gerados com o programa PRIAM, para cada um destes com mais de uma sequência, o programa MUSCLE [28] é utilizado para gerar um alinhamento múltiplo das sequências, esse alinhamento é então passado para o programa HMMBuild. Com os perfis preparados, a entrada do método consiste em um conjunto de sequências de DNA, sejam elas de cromossomos, *contigs* ou *Expressed Sequence Tags*. Uma primeira busca é feita utilizando PSI-TBLASTN para detectar sequências de DNA com alguma similaridade com algum perfil PRIAM, sequências com alguma similaridade são então passadas para o método Wise2 para posterior análise, e sequências alinhadas aos perfis são reportadas. Essa ferramenta também inclui o módulo SHARKView, que gera uma visualização da rede metabólica gerada. Como se pode ver, essa ferramenta não se dedica somente à predição de reações, mas também oferece um serviço de visualização das informações obtidas.

A ferramenta RAST [10] é voltada para genomas bacterianos e do reino *Archea*, um exemplo da possibilidade de especializar ferramentas de criação de redes metabólicas. O princípio da ferramenta é a possibilidade de projetar regras biológicas já confirmadas com experimentos em novas sequências. Essas regras biológicas conhecidas são armazenadas em subsistemas. Um subsistema consiste em um conjunto de genes de diferentes organismos que implementam essas regras biológicas. Para a execução, a ferramenta utiliza FIGfams, uma entidade com três componentes: uma família de proteínas, uma função executada pela família, e um procedimento de decisão. O procedimento de decisão recebe uma sequência e decide se a sequência deve ser incluída na FIGfam ou não. Dois procedimentos são utilizados: no primeiro, se duas proteínas implementam a mesma função de um mesmo subsistema e se as regiões de similaridade das duas sequências cobrem pelo menos 70% de ambas, a proteína é incluída na família, no segundo, se as proteínas vêm de genomas próximos, tem similaridade maior de 90% e tem um contexto cromossômico igual, então elas são incluídas na mesma família. O método toma então como entrada arquivo

multifasta contendo *contigs*, determina a localização de genes de tRNA e rRNA e os remove. Após isso, estima os genes prováveis utilizando GLIMMER2, toma genes representativos de FIGfams universais ou quase universais em procariotos, como tRNA sintetases. Essas sequências representativas são buscadas nos genes prováveis, com isso localizando os vizinhos filogenéticos mais próximos, genes encontrados nesse processo são considerados como tendo a função determinada. O próximo passo procura as FIGfams dos vizinhos filogenéticos mais próximos no conjunto de genes sem função determinada, novamente, aqueles com similaridade são considerados determinados e removidos das próximas buscas. A próxima busca é feita utilizando todo o conjunto de FIGfams, e por fim uma última busca com um banco de proteínas não redundantes pode ser feita. Feito isso, os dados de sequências associadas a subsistemas com um conjunto de cenários pode ser utilizado para gerar uma rede de reações do organismo, que pode ser vista como uma rede metabólica.

O Pathway Tools [45] é uma suíte de *software*, centrada em uma espécie de banco de dados conhecido como banco de dados de vias/genomas (BDVG). Um BDVG integra informações sobre genes, proteínas, rede metabólica e regulatória de um organismo. A criação de um BDVG começa a partir de um genoma anotado, para isso é utilizado um conjunto de arquivos fasta, cada um contendo a sequência de um elemento¹, com isso obtém-se a sequência dos elementos. A anotação é informada em arquivos no formato PathoLogic ou Genbank. A anotação pode conter informações de início e fim de genes e *introns*, pode também conter termos *Gene Ontology* (GO), *EC numbers*, descrição do produto do gene, nome do gene, nome do produto do gene, *links* para outros bancos de dados e comentários. O Pathway Tools então copia alguns dados do banco de dados MetaCyc e cria um novo objeto para cada elemento anotado dado como entrada, o objetivo deste passo é transferir as informações de sequência e anotação para o BDVG.

A seguir, o Pathway Tools utiliza os produtos de genes, os *EC numbers* e termos GO para identificar as reações do MetaCyc correspondentes às anotações das sequências, dessa forma criando um conjunto de reações. Em seguida, o Pathway Tools importa todas as vias do MetaCyc contendo pelo menos uma reação identificada no organismo, e faz uma triagem dessas vias para tentar remover falsos positivos. Utiliza-se como critério o número de reações na via e também quantas das reações na via são exclusivas daquela via e não ocorrem em outra via metabólica. Adicionalmente, uma filtragem taxonômica também é feita. Cada via tem um escopo taxonômico onde se espera a ocorrência da via, esse escopo é construído baseado em informações experimentais. Assim, vias só verificadas em plantas são eliminadas de um BDVG de um animal.

As vias identificadas podem conter falhas. Uma possível falha consiste na não identificação de uma enzima que catalisa uma ou mais reações na via. Dessa forma, ocorre a falta de uma ou mais reações na via, sendo essas reações faltantes chamadas de buracos. O Pathway Tools possui uma ferramenta para encontrar candidatos para preencher esses buracos, chamada PHFiller. Para cada buraco a ser processado, a ferramenta busca no banco de dados UNIPROT [7] todas as sequências que catalisam essa reação em outros organismos. Um BLAST é feito para

¹cromossomo, plasmídeo ou *contig*

comparar as proteínas do organismo com o conjunto de sequências do UNIPROT. A análise do resultado do BLAST é feita utilizando um classificador Bayesiano, que retorna a probabilidade de dada sequência codificar a enzima que catalisa a reação faltante.

O Pathway Tools permite a edição das informações do BDVG. Um BDVG pode ter as informações publicadas na internet, e versões do BDVG podem ser salvas para permitir a recuperação do estado anterior de um BDVG se isso for necessário.

2.3.3 Redes metabólicas reconstruídas

Várias redes metabólicas já foram reconstruídas *in silico*. Ma e Zeng [52] reconstruíram redes metabólicas *in silico* de 80 diferentes organismos, e apresentaram várias análises sobre a representação em forma de grafos dessas redes, por exemplo distribuição do grau de conectividade, para verificar a “proximidade” entre os compostos. Para organismos modelo, são utilizados métodos mais elaborados, onde uma extensa verificação de informações é feita para que a rede metabólica reconstruída represente da forma mais fiel possível o metabolismo do organismo de estudo. Como exemplo, temos a reconstrução da rede metabólica da planta modelo *Arabidopsis thaliana* [62].

A bactéria Gram-negativa *Escherichia coli* é um micro-organismo muito estudado, com vários dados sobre metabolismo, regulação e fisiologia. Ao longo do tempo, vários diferentes modelos sobre o metabolismo de *E. coli* foram gerados, antes mesmo da disponibilidade do genoma. Em 1990, construiu-se um FBA com 14 reações de *E. coli*, utilizado para prever a excreção do excesso de acetato em *E. coli* [54]. Em 1993, um modelo com 53 reações catabólicas e 94 reações biosintéticas foi criado, e dados como produção ótima de cofatores foram calculados [91]. Esse modelo foi posteriormente expandido para abranger 300 reações e 289 metabólitos, e modificações foram feitas para relacionar biomassa com a taxa de crescimento [70]. Esse modelo foi usado para calcular o fluxo máximo quando *E. coli* crescia em um meio com glicose e acetato e outro meio apenas com acetato. A diferença média entre o valor previsto e o valor verificado experimentalmente foi de 16% para o meio com glicose e acetato e 17% para o meio somente com acetato. Em 1997, o genoma de *E. coli* MG1655 foi sequenciado e anotado [17]. Em 2000, um novo modelo foi construído, com 720 reações e 436 metabólitos. Estudos de remoções *in silico* de genes de *E. coli* foram feitas com esse modelo, sendo 68 consistentes com observações experimentais das 79 previsões feitas. Outros modelos foram gerados desde então, cada um expandindo o conhecimento de modelos anteriores, por exemplo os modelos de 2007 [30] e 2011 [67].

A levedura *Saccharomyces cerevisiae* foi o primeiro eucarioto a ter a rede metabólica reconstruída *in silico* em escala genômica [35]. Outros modelos foram construídos posteriormente [27, 49], mas variações nos métodos usados e diferentes interpretações da literatura acarretaram grandes diferenças entre as reconstruções. Em 2007, foi feita uma reunião com especialistas de diversas áreas com o intuito de resolver as diferenças entre os diferentes modelos, remover as ambiguidades e gerar uma rede metabólica consenso [39]. Um dos problemas dessa rede metabólica consenso foi a fragmentação, pois reações com ambiguidades estruturais foram re-

movidas. Uma atualização desse modelo foi feita em 2010, com melhor representação do metabolismo de lipídios e maior conectividade, consistindo essa reconstrução de 1.102 reações e 924 metabólitos [26]. Mais uma atualização foi feita em 2012, atualizando reações envolvidas em esfingolipídeos, anotações gene-reação, e outras melhoras, esse novo modelo contém 2.110 reações e 1.418 metabólitos [38].

Capítulo 3

Paracoccidioides lutzii

Este capítulo apresenta o organismo para o qual propomos uma primeira visão da rede metabólica, o *Paracoccidioides lutzii*. Na seção 3.1, uma introdução geral aos fungos é feita, com aspectos biológicos e o papel destes organismos na natureza e sociedade humana. Em seguida, na seção 3.2 descrevemos o fungo, com alguns aspectos biológicos gerais, além de mais detalhes sobre a doença causada pelo fungo. Após isso, na seção 3.3, o Projeto Genoma Pb é apresentado, sendo descrita de forma resumida a sua bioinformática e os resultados obtidos.

3.1 Descrição geral de fungos

Fungo é o nome dado a qualquer organismo pertencente ao reino *Fungi*, este reino surgiu há 570–900 milhões de anos atrás e é reconhecido como um grupo independente de organismos assim como as plantas e os animais. Com pelo menos 70.000 espécies descritas, os fungos variam desde organismos unicelulares até colônias multicelulares, são heterotróficos e encontrados em diversos *habitats*.

Fungos exercem um importante papel na natureza e na sociedade humana. Fungos estão ativamente envolvidos em diversos ciclos de matéria na natureza. Como exemplo, alguns fungos são capazes de degradar a celulose, envolvidos assim no ciclo do carbono, outros fungos são capazes de fixar nitrogênio da atmosfera, participam assim do ciclo do nitrogênio. Essa capacidade de fixação de nitrogênio tem um importante papel para as plantas, pois consiste na principal forma de obtenção de nitrogênio, de fato, algumas plantas e fungos vivem em simbiose, obtendo assim mútuo proveito. Fungos associados a insetos permitem aos insetos habitar novos locais, ao prover enzimas capazes de digerir a matéria orgânica presente em um novo *habitat*. O parasitismo também é observado entre os fungos, sendo capazes de infectar plantas como o arroz, animais como os morcegos e até mesmo criaturas marinhas. Fungos são capazes de gerar esporos, esses esporos podem ser carregados pelo vento por grandes distâncias, sendo capazes de atingir outros continentes, contribuindo para a dispersão dos fungos por grande parte do planeta.

Na sociedade humana, os fungos podem ser utilizados para diversas finalidades, econômicas, médicas, culinárias, entre outras. A *Saccharomyces cerevisiae* é utilizada como fermento na produção de pão e bebidas alcoólicas, *Tolyocladium inflatum* é utilizado na produção de drogas imunossupressoras e *Penicilium chryso-*

genum, o produtor da penicilina, teve um grande impacto no progresso médico, além disso, vários fungos são utilizados na alimentação. Assim como na natureza, fungos também são capazes de parasitar seres humanos, vencendo as barreiras impostas pelo sistema imune e proliferando dentro do organismo ou até mesmo dentro das células. Exemplos de fungos capazes de infectar humanos incluem, mas não estão limitados a: *Candida albicans*, *Aspergillus fumigatus*, *Histoplasma capsulatum* e *Cryptococcus neoformans*. Infecções sistêmicas desses fungos podem levar à morte, sendo, em alguns casos, a causa de morte em ambientes hospitalares, onde após doenças ou cirurgias, pacientes são infectados por fungos.

Os fungos possuem mais de uma opção para a reprodução, reprodução sexuada ou assexuada, no entanto, existem espécies capazes apenas de reprodução sexuada, espécies que efetuam apenas reprodução assexuada também foram observadas. De maneira geral, a reprodução gera esporos capazes de dispersão e de germinação em um substrato adequado. A dispersão pode se dar de diversas formas, pelo vento, na água, através de erupção, por meio de insetos, entre outros. Certas espécies de fungos com reprodução sexuada são capazes de efetuar a auto-fecundação, enquanto outras necessitam de dois indivíduos com formas sexuais diferentes para a reprodução.

Existem duas formas celulares principais para o nosso interesse, micélios ou filamentos, em geral ramificados e chamados de hifas, e a forma de levedura. As hifas são constituídas de diversas células, estas células podem ser separadas umas das outras por paredes celulares, ou a hifa pode não exibir uma compartimentalização das células. A forma de levedura é mais simples, constituindo de uma única célula cuja multiplicação ocorre por brotamento. Existem espécies de fungos com a capacidade de assumir as duas formas celulares, fungos chamados dimórficos, bem como fungos restritos a apenas uma das formas.

3.2 Descrição do *Paracoccidioides lutzii*

Nesta seção inicialmente descrevemos o gênero *Paracoccidioides*, em seguida discutimos detalhes do *P. lutzii*

3.2.1 O gênero *Paracoccidioides*

Adolpho Lutz realizou o primeiro isolamento de fungos do gênero *Paracoccidioides*, a partir de pacientes contaminados na cidade de São Paulo no ano de 1908 [51], em 1930 foi criado o gênero *Paracoccidioides*. Fungos desse gênero são encontrados em áreas rurais da América Latina, desde o México até o extremo sul da Argentina [77], embora o *habitat* não tenha sido completamente determinado, acredita-se que eles vivam como micélios saprófitas no solo e/ou em vegetais, e usem matéria orgânica em decomposição como fonte de nutrientes [82]. As condições climáticas consistiriam em temperaturas de 10 a 28°C, vegetação com florestas abundantes, estações com verões chuvosos, invernos curtos e solos ácidos. Os tatus são reconhecidos como reservatórios naturais do fungo [11], embora ele tenha também sido encontrado em cães da raça dobermann [76] e morcegos. Autores especulam que o

reservatório natural de fungos do gênero *Paracoccidioides* consistiria em animais cujo ambiente propicia contato com a água, como peixes, anfíbios e moluscos. As aves aquáticas, após ingerirem estes animais, defecariam o fungo no solo, possibilitando a propagação do mesmo em outros ambientes [23].

O gênero *Paracoccidioides* consiste de fungos dimórficos, ocorrendo como micélio no ambiente natural e passando para a forma de levedura ao infectar o homem [75]. A forma miceliana produz colônias brancas, pequenas e irregulares, as hifas são multinucleadas e segmentadas, podendo ter esporos terminais ou intercalares. Microscopicamente a forma de levedura é multinucleada, arredondada e algumas leveduras apresentam múltiplos brotamentos, levando a um formato denominado “roda de leme”, este formato é utilizado no diagnóstico da infecção pelo fungo, a figura 3.1 é uma foto de uma célula de *P. brasiliensis* em formato “roda de leme”.

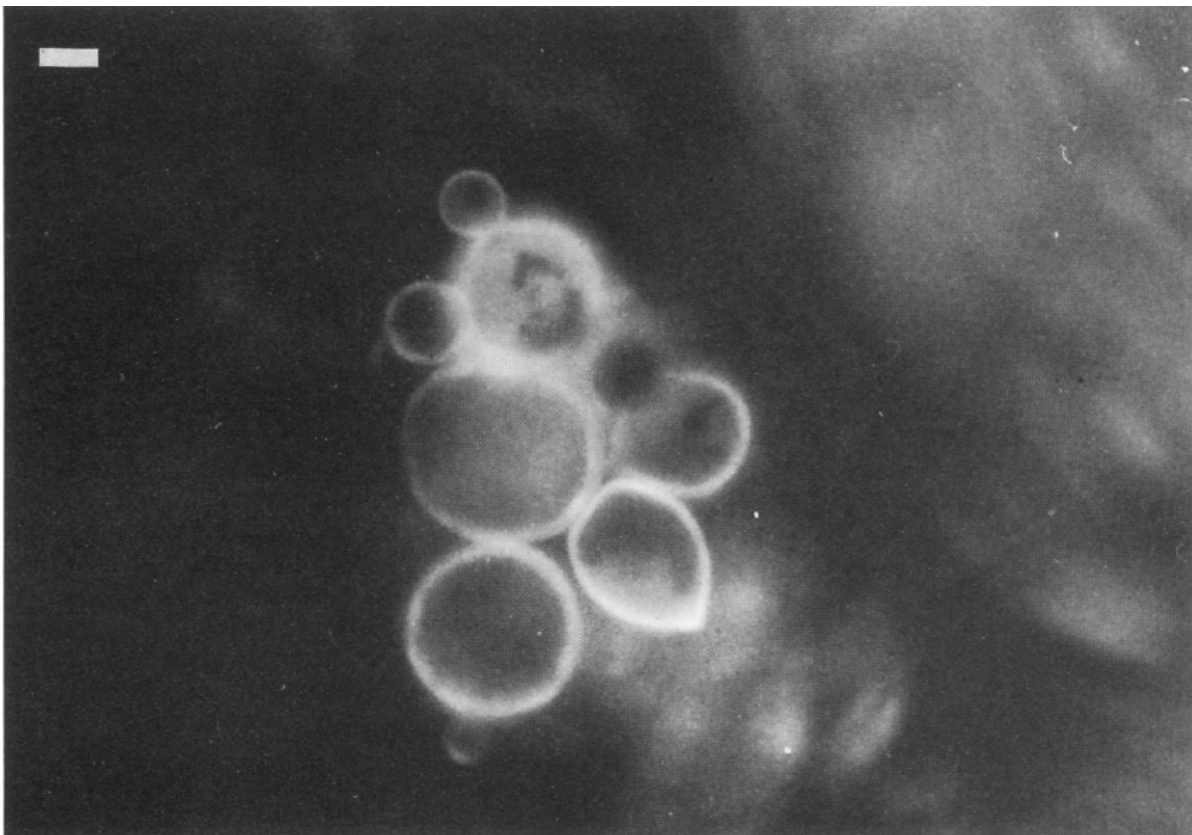


Figura 3.1: Foto de célula de *P. brasiliensis* com múltiplos brotamentos. Esse formato é conhecido como “roda de leme”. A barra branca representa $10 \mu\text{m}$. Retirado de [19].

Estudos utilizando a técnica de eletroforese de gel em campo pulsado permitiram estimar o número de cromossomos de diferentes isolados, com o número de bandas variando de 4 a 5 e tamanho estimado entre 23 e 31 milhões de pares de bases [2, 63]. Sequenciamento de fragmentos de DNA mostrou uma densidade de um gene a cada 3500–4500 mil pares de bases, para um total de 7.500–9.000 genes. O Instituto BROAD efetuou o sequenciamento do genoma de três isolados

do *Paracoccidioides*, conhecidos como Pb01, Pb03 e Pb18, os dados do sequenciamento estão disponíveis publicamente.

3.2.2 O *Paracoccidioides lutzii*

Durante muito tempo, o gênero *Paracoccidioides* teve apenas uma espécie, o *Paracoccidioides brasiliensis*, e diferentes “cepas” ou isolados. Os isolados Pb01, Pb03 e Pb18 são os mais estudados do gênero *Paracoccidioides*. Estudos utilizando o método de *Genealogical Concordance Phylogenetic Species Recognition* (GCPSR) identificaram três espécies filogenéticas no complexo *Paracoccidioides*: S1, PS2 e PS3 [59]. Um outro estudo feito com marcadores moleculares do gene *hsp70* identificou 17 isolados, incluindo o Pb01, nomeados grupo *Pb01-like*, análises feitas com GCPSR mostraram que o grupo *Pb01-like* clusterizou separadamente dos grupos S1, PS2 e PS3 em um longo ramo monofilético [88].

Análises morfológicas das conídias de espécies agrupadas como *Pb01-like* mostrou conídias alongadas com tamanho maior em comparação com as conídias de *P. brasiliensis* [90], além disso, as conídias de *P. brasiliensis* são globulares. Dados genéticos como sequências de PRP8, além de elementos transponíveis, também corroboram a existência da espécie *P. lutzii* [57, 89].

Com a disponibilidade dos genomas dos isolados Pb01, Pb03 e Pb18, estudos verificaram divergência genética entre Pb01 e Pb03, Pb18. Os isolados Pb03 e Pb18 tem genoma de tamanho similar (30 e 29,1 Mb), já o genoma do Pb01 tem 32,9 Mb [25]. Não obstante, os isolados Pb03 e Pb18 exibem uma similaridade de sequências maior (96%), e o isolado Pb01 apresenta maior divergência (90%) [25]. *P. lutzii* exibe uma localização geográfica diferente, sendo encontrado nos estados de Goiás a Rondônia, enquanto o Pb18 é encontrado nos estados da Região Sudeste.

3.2.3 Paracoccidioidomicose

Fungos do gênero *Paracoccidioides* causam uma doença conhecida como Paracoccidioidomicose, abreviada como PCM. A PCM é uma micose sistêmica com distribuição não uniforme na América Latina, com nenhum caso reportado nos seguintes países: Nicarágua, Guiana, Chile e Suriname. O Brasil concentra cerca de 80% dos casos, observados nos estados de São Paulo, Goiás, Minas Gerais e Mato Grosso. O número de pessoas afetadas pela doença é estimado em 10 milhões [19]. O mecanismo de infecção foi especulado da seguinte maneira: a forma miceliana no solo libera propágulos, esses são então dispersos pelo ar e ao serem inalados poderiam provocar a doença [75]. Experimentos onde camundongos inalam esses propágulos verificaram a instalação de uma doença crônica e progressiva nos camundongos [60], podendo induzir fibrose pulmonar, como verificado em humanos.

A doença apresenta diversos quadros clínicos, a depender do parasita e do hospedeiro, e apresenta um tempo de latência alto, com uma média de 15 anos. A infecção pode originar uma forma assintomática da doença, com uma infecção primária no tecido pulmonar. As formas infectantes posteriormente assumem a morfologia de levedura, sendo este passo essencial para o prosseguimento da doença, cepas incapazes de assumir a forma de levedura não causam infecção. Acredita-

se que o aumento da temperatura experimentado no corpo humano ($37^{\circ}C$), tenha um papel fundamental na diferenciação, pois experimentos em laboratório envolvendo a variação de temperatura provocaram a transição entre as duas formas do fungo [61]. A figura 3.2 as duas formas de micélio e levedura do fungo.



Figura 3.2: As duas formas do *P. lutzii*, micélio (a direita) e levedura (a esquerda)

Após a instalação existem duas formas básicas de apresentação da doença: aguda e crônica. A forma aguda afeta ambos os sexos, tem evolução rápida, na escala de tempo de semanas a meses, a mortalidade pode ser severa pois esta forma da doença pode se espalhar para baço, fígado e medula óssea [19]. A forma crônica ocorre preferencialmente em homens adultos, levando anos para apresentação de sintomas, e afeta principalmente o endotélio pulmonar, podendo posteriormente se espalhar para outros órgãos [36]. A preferência por homens pode ser explicada por uma diferença hormonal: foi demonstrado um poder inibitório do hormônio β -estradiol na mudança de forma do fungo [78], dessa forma mulheres teriam uma proteção natural contra a infecção pelo fungo.

O diagnóstico da doença se dá principalmente por meio da identificação direta do fungo em material retirado das lesões, procurando principalmente a “roda de leme” característica. O tratamento é prolongado, com possibilidade de chegar a 5 anos. Utilizam-se principalmente sulfas e drogas anti-fúngicas [58], com posterior acompanhamento pós-terapêutico, devido ao grande número de infectados pela doença, estudos envolvendo a busca por vacinas ou outro tratamento mais eficaz foram efetuados.

3.3 Projeto Genoma Pb

O Projeto Genoma Pb foi criado com o seguinte objetivo geral:

Mapear o genoma funcional e diferencial entre as formas de micélio e levedura de *Paracoccidioides brasiliensis*, um fungo dimórfico que causa a Paracoccidioidomicose (PCM), micose endêmica de alta prevalência na América Latina.

As coordenações geral e da bioinformática ficaram sob responsabilidade do Instituto de Biologia da UnB, e envolveram diversas instituições de pesquisa desta

região, como a Universidade Federal de Goiás, a Universidade Federal de Mato Grosso do Sul e a Universidade Federal de Mato Grosso, entre outras.

Na etapa de sequenciamento foram gerados 25.598 clones de cDNA e estes clones foram sequenciados. A determinação das bases foi feita utilizando o programa phred [29], logo após seguiu-se uma remoção de sequências contaminantes de bactéria e rRNA, por fim foram aceitas para montagem somente sequências com 200 ou mais bases e qualidade PHRED ≥ 20 . Esse processo gerou 19718 sequências de alta qualidade [32], essas sequências foram montadas utilizando o programa CAP3 [40], gerando 2.655 *contigs* e 3.367 *singlets*. Posteriormente *contigs* e *singlets* foram anotados utilizando-se blastx [3] e os bancos de dados *nonredundant protein* (nr), *cluster of orthologous groups* (COG) [48] e *gene ontology* (GO) [9]. A figura 3.3 mostra de maneira simplificada o *pipeline* de bioinformática utilizado no projeto.

De forma geral, 27 genes altamente transcritos foram encontrados nesse transcrito, 328 genes tiveram expressão aumentada no processo de diferenciação, sendo 58 de micélio e 270 de levedura. Alguns padrões de expressão diferenciada de genes foram validados experimentalmente por meio da técnica de *northern blot*. Também foram indicados alguns ECs durante o processo, a tabela 3.1 apresenta alguns dos ECs identificados como diferencialmente expressos e a forma do fungo onde eles são mais expressos [32].

Tabela 3.1: Tabela de ECs diferencialmente expressos identificados no Projeto Genoma Pb. Dados obtidos de [32].

ECs encontrados	
Levedura	Micélio
1.1.1.1	4.4.1.5
3.5.1.41	1.15.1.1
3.1.6.6	1.13.11.32
1.8.4.8	2.7.4.3
1.2.4.1	4.2.1.22
2.6.1.42	2.2.1.2
1.6.5.3	3.1.2.22
1.1.1.69	1.1.1.41
3.3.2.1	6.2.1.5
2.5.1.15	2.7.1.2
2.6.1.1	2.7.1.48
6.2.1.3	
6.3.5.7	
4.1.3.1	
2.6.1.9	
3.5.1.4	
1.14.13.3	

Uma análise dos genes permitiu constatar a capacidade de respiração aeróbica por parte do fungo bem como a capacidade de respiração anaeróbica pela conversão de piruvato para etanol. Um estudo do padrão diferenciado de expressão leva a crer

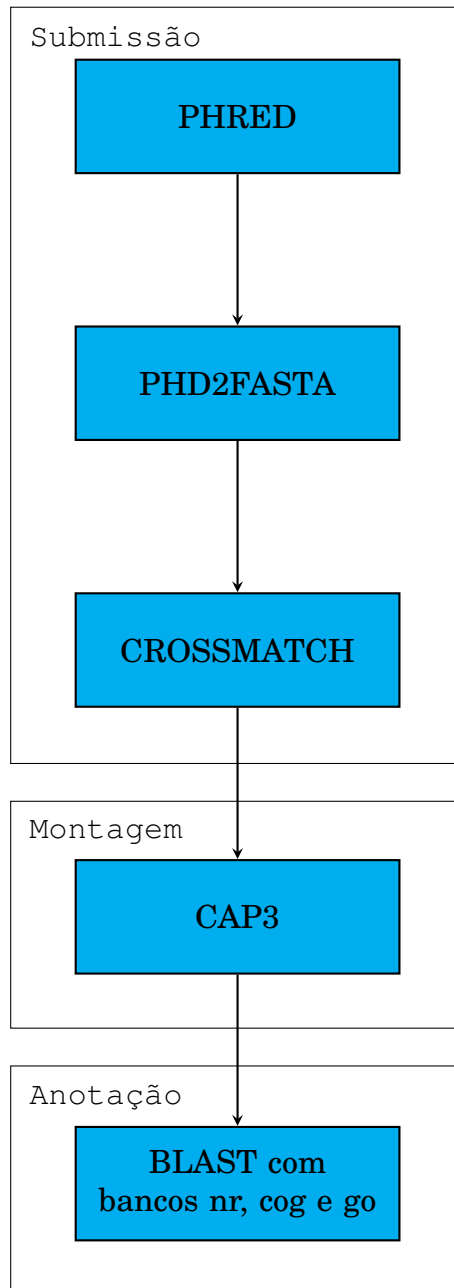


Figura 3.3: *Pipeline* de bioinformática do Projeto Genoma Pb.

que a forma de micélio utiliza principalmente a respiração aeróbica, já a forma de levedura tem preferência por respiração anaeróbica.

Em Arraes et al. [8], um estudo geral sobre o metabolismo de *P. lutzii* é feito, abordando tanto o metabolismo primário quanto o secundário, além de serem apresentadas referências para experimentos que confirmam alguns aspectos da biologia do fungo. Nosso trabalho também utiliza esses dados, além de outros encontrados na literatura e bibliografia específica para a reconstrução da rede metabólica.

Capítulo 4

Métodos

Neste capítulo, descrevemos inicialmente o método geral utilizado para a reconstrução *in silico* da rede metabólica do *P. lutzii* e em seguida os métodos computacionais utilizados para adquirir e analisar os dados. Na seção 4.1 será feita uma descrição do método geral. Na seção 4.2 apresentamos os métodos computacionais utilizados no desenvolvimento do trabalho.

4.1 Descrição do método geral

Baseado no trabalho de Forster et al.[35], várias fontes de dados foram utilizadas como entrada para diferentes parâmetros da rede metabólica *in silico*. O genoma obtido pelo Instituto BROAD [Broad Institute of Harvard and MIT] além da anotação feita com o KAAS [64] foram utilizados para a procura de enzimas. Essas enzimas foram utilizadas para identificar as reações e respectivas vias metabólicas. Além disso, artigos sobre o *P. lutzii* serviram de base para incluir ou confirmar enzimas que deveriam estar presentes [8, 32, 33]. O banco de dados MetaCyc foi utilizado para obter informações como estequiometria e cofatores de enzimas, além de conter os dados das reações propriamente ditas. Um livro de bioquímica básica foi consultado para uma maior compreensão dos dados analisados. Por fim, todas essas informações foram incluídas para propor um modelo *in silico* da rede metabólica do organismo. A figura 4.1 apresenta de forma resumida esse método.

O *software* Pathway Tools [45] foi sugerido por pesquisadores do instituto INRIA da França, durante um período de pesquisa de setembro a dezembro de 2012. O Pathway Tools vem sendo desenvolvido desde 2000, e é a base de diversos bancos de dados tais como o do projeto Ecocyc. Além disso, permite a atualização dos dados da rede metabólica conforme novos conhecimentos venham a ser descobertos. A edição pode ser feita via *web* e de maneira colaborativa, assim pode-se deixar a atualização de informações a cargo de um especialista. É possível também definir versões do banco e guardar essas versões separadamente, permitindo restaurar os dados a um estado antigo mais facilmente. O programa também permite realizar uma série de outras análises, tais como *Flux Balance Analysis* (FBA), e identificação de metabólitos nunca produzidos ou jamais consumidos.

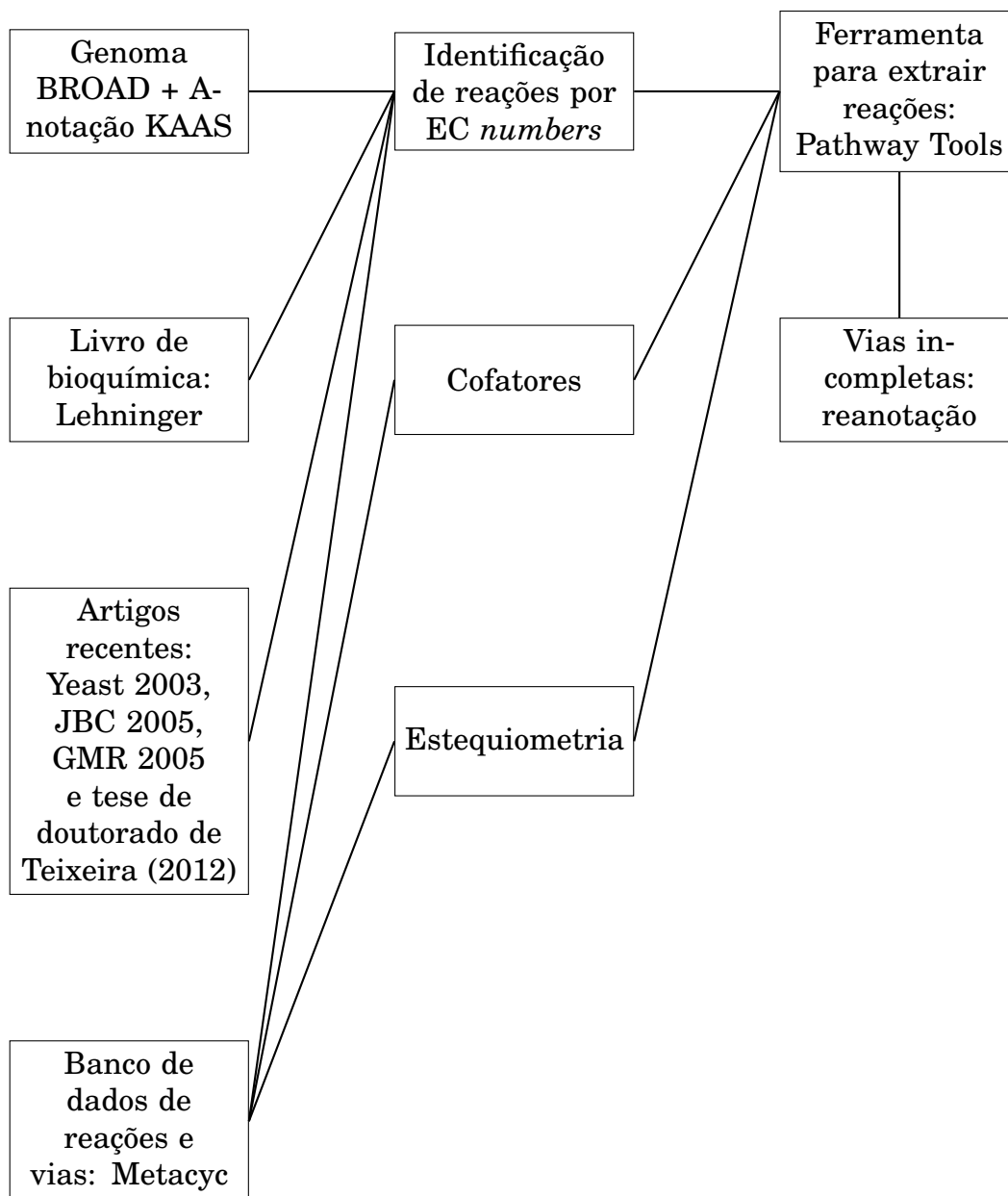


Figura 4.1: Método geral de reconstrução da rede metabólica.

4.2 Descrição do método computacional

Nesta seção apresentamos os processos de identificação das enzimas do *P. lutzii*, bem como a obtenção e a modificação nos arquivos de vias provenientes do KEGG e do Biocyc. Por fim, descrevemos a geração de dados de entrada para a execução do Pathway Tools de modo a reconstruir a rede metabólica *in silico* do fungo.

4.2.1 Identificação de enzimas

Primeiramente, obtivemos os dados do Instituto Broad referentes ao *Paracoccidioides lutzii*, identificado no Broad como *P. brasiliensis* Pb01, a partir do endereço http://www.broadinstitute.org/annotation/genome/paracoccidioides_brasiliensis/MultiDownloads.html. Os dados obtidos consistem no agregado *Gene Downloads* contendo os seguintes arquivos:

- genes.fasta
- transcripts.fasta
- transcripts.gtf
- proteins.fasta
- proteins_stops.fasta
- pfam_to_genes.txt
- genes_upstream_1000.fasta
- genes_upstream_utr_1000.fasta
- genes_downstream_1000.fasta
- genes_downstream_utr_1000.fasta
- genome_summary.txt
- genome_summary_per_gene.txt

Após isso, efetuamos um Blast [3] do arquivo proteins.fasta contra o banco de dados swissprot, onde todas as comparações com *e-value* $> 10^{-10}$ foram rejeitados, sendo o resultado deste blast salvo em um arquivo xml para posterior processamento. Em seguida, utilizamos um *script* python para obter os identificadores dos melhores *hits* de cada sequência contra o swissprot, sequências sem *hits* não foram consideradas nesta etapa. Os identificadores foram salvos em um arquivo contendo um identificador por linha. Acessamos o endereço <http://www.uniprot.org/>, clicamos na aba superior com o texto *retrieve*, carregamos o arquivo com identificadores gerado anteriormente e submetemos a transação. Obtivemos depois um arquivo flat file contendo informações de cada um dos *hits*. Utilizamos um segundo *script* python para processar o resultado do blast com os dados dos flat files obtidos do swissprot, associando as sequências com os ECs correspondentes, quando estes existiam, o *script* python grava os ECs encontrados em um arquivo texto, um EC por linha.

A figura 4.2 mostra o *pipeline* utilizado para obtenção das vias.

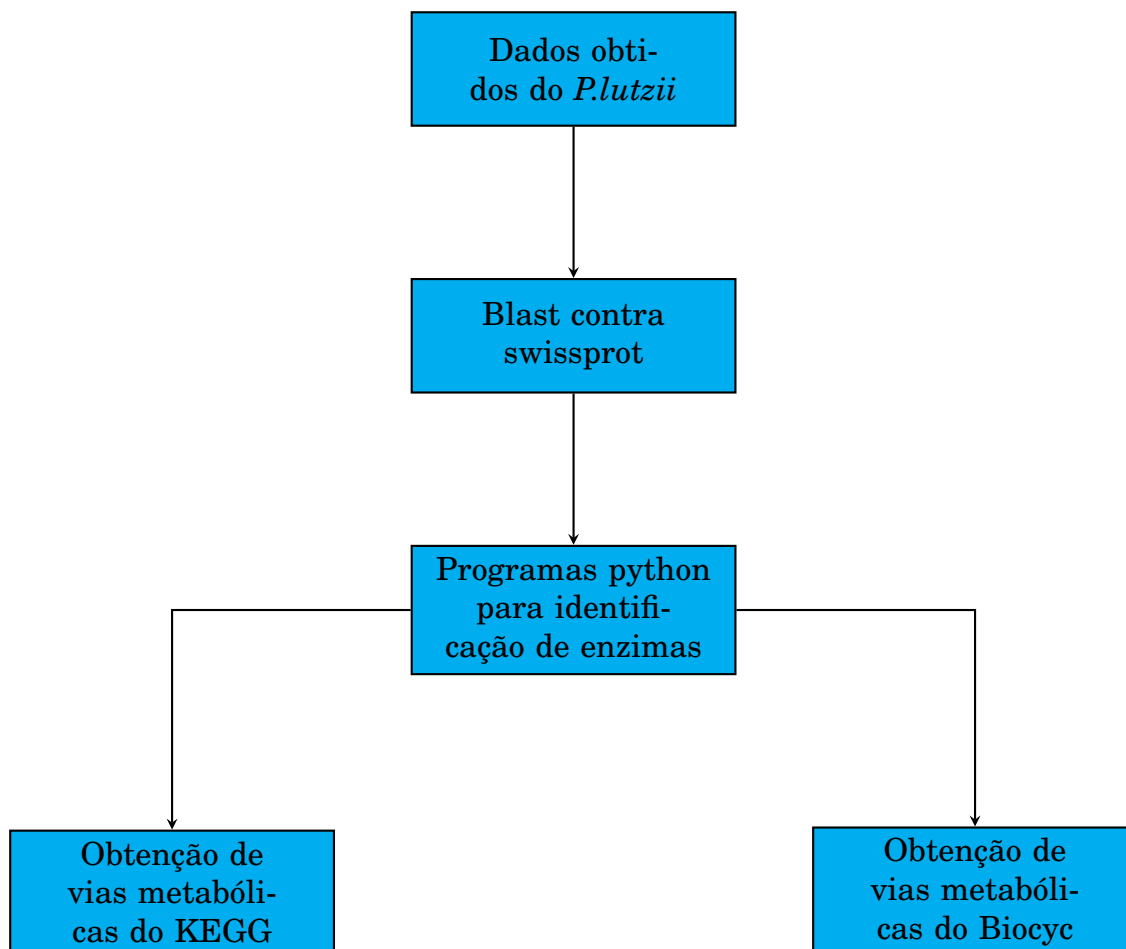


Figura 4.2: *Pipeline* para a obtenção de vias metabólicas.

4.2.2 Obtenção de vias metabólicas

Antes de prosseguir com a criação da rede metabólica, decidimos utilizar os ECs obtidos na seção 4.2.1 para obter vias metabólicas de bancos de dados públicos. O propósito era obter dados primários sobre o metabolismo.

Para esta etapa foram utilizados os dados do Kyoto Encyclopedia of Genes and Genomes (KEGG) bem como dois bancos de dados do projeto Biocyc, o Ecocyc e o Yeastcyc. Os procedimentos para o KEGG são diferentes daqueles utilizados para os dados do projeto Biocyc, descrevemos cada um desses passos em ordem.

Obtenção de vias metabólicas do KEGG

Como dito na seção 2.2.1, o KEGG permite o envio de um arquivo contendo um EC por linha para uma busca nas vias metabólicas. Um detalhe é que para ser considerado válido, cada EC deve ser precedido do texto `ec:`, isto é, o texto `ec:2.1.7.1` é válido para o KEGG, já o texto `2.1.7.1` não é válido. Utilizamos um *script* para gerar o arquivo no formato utilizado pelo KEGG. Acessamos o endereço http://www.genome.jp/kegg/tool/map_pathway1.html, carregamos o arquivo com os ECs no formato do KEGG, escolhemos um organismo de interesse (os organismos utilizados serão detalhados depois) e efetuamos a busca.

O KEGG retorna uma página de resultados onde se encontra uma lista de *links*, cada *link* aponta para uma página onde se pode visualizar a via metabólica com os destaques apropriados. Infelizmente essas páginas são apagadas após um tempo, dessa forma, para tornar o resultado persistente, é necessário salvar as páginas de resultado. Salvamos o arquivo html descrevendo a página contendo a lista de resultados, utilizamos um *script* python para varrer essa lista de resultados e processá-los, da seguinte forma: para cada *link* de resultado, a página correspondente é obtida, o arquivo contendo a imagem da via com os destaques é salvo em um diretório local, também é salvo um pequeno código html cuja função é permitir aos dados da imagem serem utilizados como *links* para mais informações no KEGG.

O processo descrito foi executado para os seguintes organismos: *Escherichia coli*, *Saccharomyces cerevisiae*, *Coccidioides immitis*, *Aspergillus fumigatus* e *Aspergillus nidulans*.

Obtenção dos dados do projeto Biocyc

Os bancos de dados do projeto Biocyc não permitem a busca de um conjunto de ECs, cada EC deve ser buscado manualmente no banco de dado. O resultado da busca não necessariamente aponta para uma via metabólica, mas pode fornecer zero, um ou mais *links* para páginas de reações, cada uma dessas páginas de reações pode também conter zero, um ou mais *links* para páginas de vias, e as páginas de vias podem vir por padrão em diferentes níveis de detalhes, algumas mostram os ECs da via, outras não mostram os ECs da via, mas o nível de detalhes pode ser ajustado.

Os endereços utilizados para busca das vias implementam proteção contra *web crawling*. *Web crawling* consiste na prática de utilizar um programa de computador automático para baixar uma grande quantidade de páginas de um endereço específico. Ele pode ser utilizado para, por exemplo, replicar as páginas públicas

de um endereço em um computador pessoal. Infelizmente, o uso de *web crawling* tem um impacto negativo no endereço cujos dados são copiados, devido ao grande número de requisições, a cota de banda pode ser excedida. Além disso, tal prática prejudica o acesso de outros usuários, e portanto é considerada indesejável pelos responsáveis dos serviços hospedados na internet. Normalmente a proteção contra *web crawling* é implementada bloqueando o acesso de computadores com muitas requisições em um curto período de tempo. A proteção contra *web crawling* implementada de maneira agressiva, por exemplo, bloquear todos os computadores acessando dois *links* em menos de 10 segundos pode excluir usuários legítimos, e isso é indesejável. Alguns programas, como o *wget*, tentam “simular” um comportamento humano obtendo os dados das páginas em intervalos de tempo aleatórios. Decidimos então utilizar a mesma abordagem, em todos os passos onde dados são obtidos das páginas do projeto Biocyc, um intervalo de tempo aleatório é introduzido antes de cada requisição, para evitar a proteção contra *web crawling*.

Por causa do método de busca, as imagens das vias não fazem distinção entre os ECs encontrados e os ECs não encontrados, sendo assim é necessário utilizar uma biblioteca de processamento de imagens para destacar os ECs encontrados. O processo para obter as vias de um banco de dados do projeto Biocyc não é essencialmente complicado, mas contém muitos passos e pode ser confuso. Por essa razão apresentamos esse processo em uma série de passos, o programa utilizado não necessariamente efetua esses passos na ordem ou na forma exposta a seguir.

Obtenção das páginas de reações

O primeiro passo consiste em enviar uma requisição contendo o EC de interesse à página do projeto Biocyc onde se busca as vias e obter o resultado. O processamento do resultado consiste em primeiro se procurar por diretivas `redirect` cuja função é apontar para uma nova página. Caso uma diretiva `redirect` seja encontrada, a página apontada é obtida, após isso processa-se a página obtida, procurando-se na mesma *links* para páginas de reações, esses *links* podem ser buscados utilizando-se uma expressão regular. Nesse processo, construímos uma estrutura de dados associando a cada EC as reações catalisadas pelo EC.

Obtenção das páginas de vias

Retira-se de cada página de reação os *links* para as vias onde essa reação ocorre, esses *links* são ajustados para o nível de detalhe necessário para visualização dos ECs nas vias e posteriormente armazenados em um conjunto para evitar duplicação de informações. Posteriormente essas páginas são baixadas localmente.

Processamento das páginas de vias

Com as páginas baixadas, é hora de processá-las para obtenção dos dados necessários. O título da via é obtido para facilitar a identificação. A imagem da via é baixada junto com as informações necessárias para *links* entre a imagem e os dados do biocyc. A imagem da via é armazenada em um arquivo local e modificada da seguinte

forma: a partir das informações dos *links* de reações em cada via, é possível descobrir os ECs associados com cada via, com essas informações é possível procurar nos *links* associados com cada imagem *links* correspondentes aos ECs buscados. Com os dados de área associado a cada *link* um quadrado vermelho é desenhado ao redor de cada área, destacando assim os ECs encontrados. Os dados da página de via são guardados em um arquivo html apontando para a imagem modificada.

Efetuamos tal procedimento para dois bancos de dados do projeto Biocyc, ambos da camada 1, isto é, bancos com uma extensa curagem manual por especialistas no organismo. Usamos os seguintes bancos: Ecocyc, de *Escherichia coli* e Yeastcyc, de *Saccharomyces cerevisiae*.

4.2.3 Construção da rede metabólica

Para a construção da rede metabólica foi utilizado o programa Pathway Tools. Alguns procedimentos foram modificados para utilização desse programa. Detalhamos os passos. Para a construção utilizamos o arquivo transcripts.fasta obtido do BROAD.

Obtenção de enzimas

O KEGG Automatic Annotation Server (KAAS) [64] oferece uma anotação de um conjunto de genes por meio de uma comparação utilizando blast [3] contra o banco de dados curado KEGG GENES. Os resultados retornados são identificadores de KEGG Orthology (KO), grupos de genes ortólogos, estes KO são associados a ECs, então é possível obter informações de enzima desta forma.

Acessamos o endereço http://www.genome.jp/kaas-bin/kaas_main, carregamos o arquivo transcripts.fasta, e escolhemos a opção de um arquivo de nucleotídeos. O conjunto de dados de genes foi escolhido como o representativo para eucariotos e o método de atribuição foi o *bi-directional best hit*. Com a análise completa baixamos um arquivo texto contendo campos separados por tabulação. O primeiro campo é o identificador da sequência, o segundo, quando presente, consistia em um identificador KO. Usamos um *script* para armazenar dados de KOs localmente e analisar esses dados para obter uma lista de ECs.

Geração dos arquivos de entrada

O Pathway Tools toma uma variedade de arquivos de entrada. O arquivo de entrada mestre, por assim dizer, é chamado genetic-elements.dat, exatamente dessa forma. O arquivo consiste em registros terminados por //, incluindo o último registro do arquivo. Cada campo do registro aparece em uma linha, e um campo é separado do valor por uma tabulação. Descrevemos brevemente os principais campos de cada registro, com alguns detalhes omitidos:

ID Obrigatório, identificador único para o elemento. Utilizamos o identificador das sequências do BROAD.

NAME Opcional, uma descrição do elemento.

TYPE Opcional, um entre :CHRSM (cromossomo), :PLASMID (plasmídio), :MT (cromossomo mitocondrial), :PT (cromossomo de cloroplasto) ou :CONTIG.

CIRCULAR? Obrigatório exceto para elementos do tipo cromossomo, indica se o elemento é circular ou não.

CODON-TABLE Opcional, o código genético a ser utilizado na tradução.

ANNOT-FILE Obrigatório, caminho completo para o arquivo em formato pathologic com a anotação do elemento.

SEQ-FILE Obrigatório, caminho completo para o arquivo fasta com a sequência do elemento.

Como dito na descrição do campo ANNOT-FILE, é necessário gerar arquivos no formato pathologic com a anotação de cada elemento. O formato desse arquivo é o mesmo do arquivo genetic-elements.dat, mas utiliza campos diferentes. Descrevemos os campos utilizados neste trabalho:

ID Opcional, identificador único usado para o elemento.

NAME Obrigatório, nome do elemento.

PRODUCT-TYPE Obrigatório, indica o tipo de produto, e deve ser escolhido entre: P (proteína), PSEUDO = pseudogene, TRNA = RNA transportador, RRNA = RNA ribossomal, MISC-RNA = outro RNA.

STARTBASE Opcional, base inicial de transcrição.

ENDBASE Opcional, base final de transcrição.

FUNCTION Obrigatório, a função do produto. Caso a função seja desconhecida, deve-se utilizar o texto ORF como função.

EC Opcional, o EC associado ao produto, se houver. Caso haja mais de um EC associado, utiliza-se tantos campos ECs quanto necessários.

Utilizamos um *script* para geração do arquivo genetic-elements.dat e um outro *script* para geração dos arquivos de anotação em formato pathologic. Um outro *script* separado foi utilizado para gerar os arquivos fasta com as sequências de cada elemento a partir do arquivo multifasta transcripts.fasta.

Execução do Pathway Tools

Executamos o Pathway Tools com os seguintes dados de entrada, observando-se que o banco de dados utilizado foi o MetaCyc:

- Nome do banco: Pb01cyc;
- Arquivo genetic-elements.dat, com referências para arquivos de sequências e anotações;

- Identificador taxonômico do NCBI: 121759

Foi então criado o banco de dados contendo as informações de reações e vias do *P. lutzii*, que é mostrado na página do projeto.

4.2.4 Reanotação

Após a execução do Pathway Tools, realizamos uma verificação manual em vias metabólicas tidas como essenciais para a sobrevivência do *P. lutzii*. O objetivo dessa verificação foi encontrar enzimas que faltavam para completar essas vias essenciais. A partir dessa verificação, obtivemos uma lista de ECs a serem buscados. Obtivemos as sequências dos ECs das enzimas não encontradas no *P. lutzii* nos bancos de dados BRENDA [80] e UNIPROT [7], e construímos um banco de dados com essas sequências.

Com esse banco de dados, separamos as sequências de *P. lutzii* sem ECs atribuídos pelo KAAS. Realizamos um BLAST dessas sequências contra o banco de dados das enzimas não encontradas nas vias de *P. lutzii*, retornando apenas os alinhamentos com *e-value* menor que 10^{-5} . Elaboramos um *script* Python para processar o resultado do BLAST, identificando as sequências com pelo menos um *hit*. A partir desses resultados, geramos duas planilhas: uma contendo novas sugestões de anotação para proteínas classificadas como hipotéticas, e outra contendo novas anotações para sequências previamente anotadas.

Capítulo 5

Resultados obtidos e discussão

Neste capítulo, descrevemos os resultados obtidos utilizando a metodologia descrita no capítulo 4, na seção 5.1 e, em seguida, discutimos esses resultados na seção 5.2.

5.1 Resultados

Nessa seção, inicialmente, mostramos as vias obtidas por meio de pesquisa nos bancos de dados KEGG e Biocyc. Posteriormente, mostramos uma representação gráfica da rede metabólica gerada bem como algumas estatísticas a respeito da rede. Por fim, mostramos algumas telas encontradas na página desenvolvida nesse projeto.

5.1.1 Identificação de enzimas e busca de vias nos bancos de dados KEGG e Biocyc

A metodologia de identificação de enzimas empregada permitiu obter 857 ECs para as proteínas do *Paracoccidioides lutzii*. A figura 5.1 mostra a razão entre o número de ECs e o número de ORFs, para cada categoria enzimática definida pela IUBMB. Destacamos que para a classe EC 5, a razão é 58/58, indicando que cada reação é catalisada por uma única ORF.

No caso do banco de dados KEGG, o número de vias metabólicas obtidas variou conforme o organismo de referência. A tabela 5.1 mostra o número de vias obtidas para cada um dos cinco organismos de referência utilizados. De maneira geral, quanto mais próximo o organismo de referência do *Paracoccidioides lutzii*, mais vias foram encontradas, a exceção consiste no empate entre *Escherichia coli* e *Saccharomyces cerevisiae*, nós propomos que a maior presença de dados de *Escherichia coli* no KEGG “compensou” a distância filogenética.

A figura 5.2 mostra uma das vias encontradas no KEGG utilizando como base de comparação os dados de *S. cerevisiae*. Os ECs são mostrados em caixas, onde caixas com fundo branco representam ECs que não pertencem a nenhum dos dois organismos. Caixas com fundo verde representam ECs encontrados em *S. cerevisiae*, bordas e números vermelhos representam os ECs de *P. brasiliensis*. Como

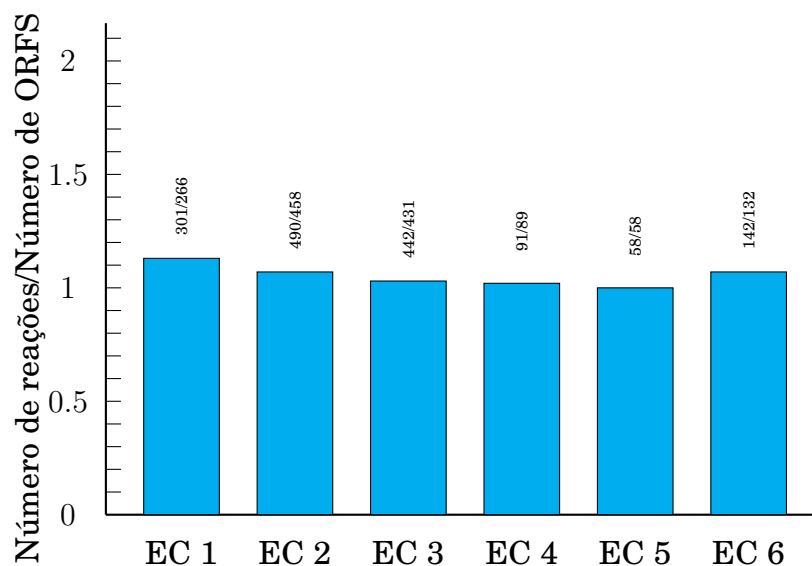


Figura 5.1: Relação entre o número de reações e o número de ORFs na reconstrução da rede metabólica, agrupados por categoria enzimática.

Tabela 5.1: Número de vias encontradas no KEGG utilizando diferentes organismos de referência.

Organismo	Número de vias metabólicas encontradas
<i>Escherichia coli</i>	96
<i>Saccharomyces cerevisiae</i>	96
<i>Coccidioides immitis</i>	101
<i>Aspergillus nidulans</i>	103
<i>Aspergillus fumigatus</i>	104

exemplo, o EC 5.4.2.2 ocorre tanto em *S. cerevisiae* quanto em *P. brasiliensis*, o EC 3.2.1.26 é exclusivo de *S. cerevisiae*.

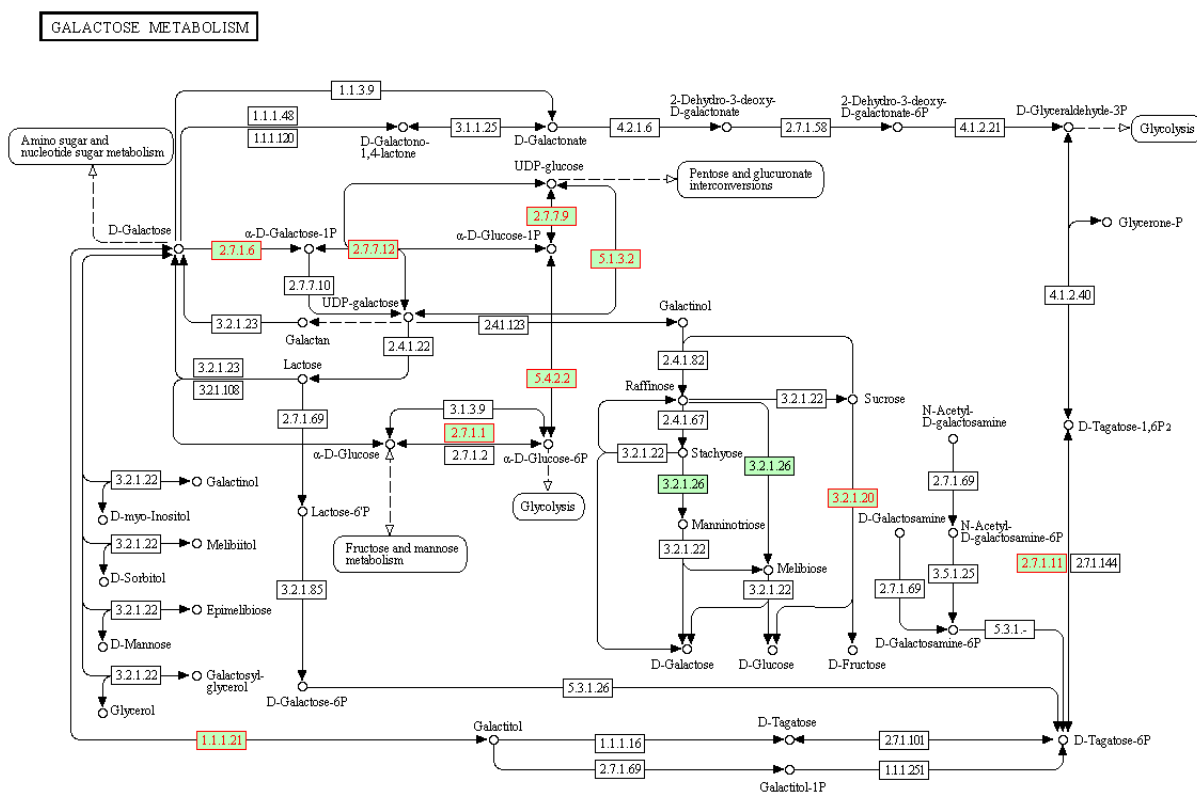


Figura 5.2: Via metabólica encontrada no banco de dados KEGG [44].

No caso dos bancos de dados do projeto Biocyc encontramos 175 vias no Ecocyc e 149 vias no Yeastcyc. A figura 5.3 mostra uma das vias encontradas no yeastcyc. O EC 4.1.1.50, destacado com uma borda vermelha, ocorre tanto em *S. cerevisiae* quanto em *P. brasiliensis*, já o EC 2.5.1.22 seria exclusivo de *S. cerevisiae*.

5.1.2 Reconstrução da rede metabólica

A rede metabólica criada possui 1743 reações catalisadas por 1413 enzimas, além de 9 reações de transporte, as reações atuam sobre um total de 1202 compostos. A figura 5.4 mostra uma visão geral da rede, sem grandes detalhes. Foram identificados 242 buracos em 106 vias, e para 166 desses buracos, pelo menos um candidato foi encontrado com a capacidade de preencher esse buraco na via metabólica.

A tabela 5.2 mostra algumas estatísticas da rede metabólica gerada.

A tabela 5.3 mostra os bancos de dados utilizados neste trabalho.

5.1.3 Página com as informações das vias metabólicas do *P. lutzii*

As figuras 5.5, 5.6, 5.7, 5.8 e 5.9 mostram informações que estão disponíveis publicamente na página: <http://www.biomol.unb.br/Pb01>.

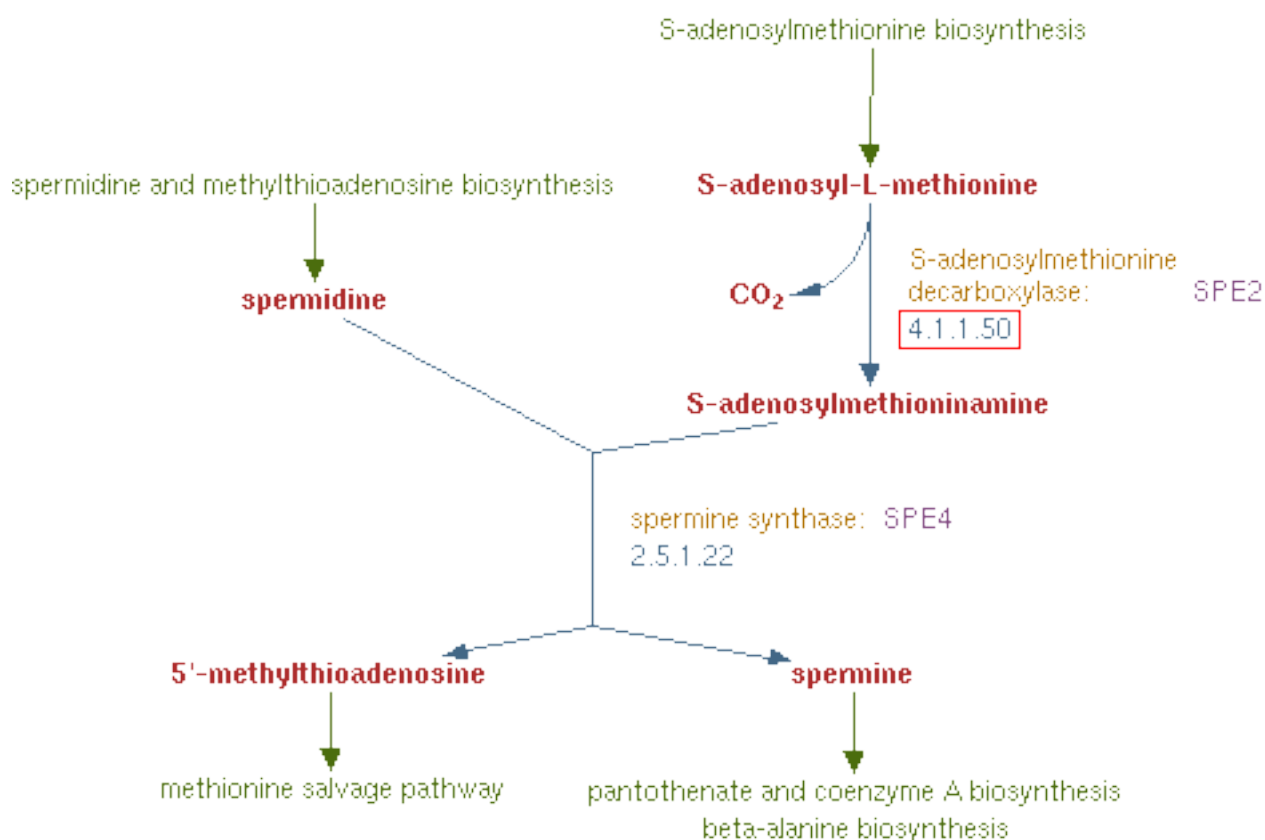


Figura 5.3: Via metabólica encontrada no banco de dados yeastcyc.

Tabela 5.2: Estatísticas da rede metabólica gerada.


Estatística	Valor
ORFs	9137
Compostos	1202
Reações	1743
Reações de transporte	9
Reações com atribuição de ORF	1555

Tabela 5.3: Bancos de dados utilizados no trabalho.


Banco	Endereço
KEGG	http://www.genome.jp/kegg/
Ecocyc	http://ecocyc.org/
Yeastcyc	http://biocyc.org/YEAST
Swissprot	http://web.expasy.org/docs/swiss-prot_guideline.html



Figura 5.4: Rede metabólica gerada.



The Pb01 Metabolic Network Project




- ▶ About *Paracoccidioides brasiliensis*
- ▶ Data sets
- ▶ Metabolic pathways
- ▶ Methodology
- ▶ Team
- ▶ Metabolic network picture


The aim of this project is to construct the metabolic network of the organism *Paracoccidioides brasiliensis* isolate 01. The *Paracoccidioides brasiliensis* (Pb) is a dimorphic fungus which causes the endemic mycosis paracoccidioidomycosis, a disease that is widespread in Latin America.

This project is being carried on at University of Brasilia (Unb) with the cooperation of researchs from Lyon, France. The University of Brasilia has already mapped the functional and differential genomes of the yeast and mycelium forms of *Paracoccidioides brasiliensis*, whose project page can be accessed [here](#). This project will built upon the data and expertise of the researchers with the ultimate goal of obtaining the Pb01 metabolic network, which can then be further researched for potential drug targets.

Figura 5.5: Página inicial



The Pb01 Metabolic Network Project




- ▶ About *Paracoccidioides brasiliensis*
- ▶ Data sets
- ▶ Metabolic pathways
- ▶ Methodology
- ▶ Team
- ▶ Metabolic network picture


Paracoccidioides brasiliensis is the etiological agent of paracoccidioidomycosis (PCM) a human systemic mycosis that mainly affects rural communities in South and Central America. PCM is manifested in different clinical forms, which are tightly dependent on the immune competence of the host. In individuals displaying the localized form of the disease it is observed the formation of granulomes and the presence of competent T lymphocytes; on the other hand, in the dispersed form of the disease, the granulomes are not formed and the T lymphocytes are hyporeactive. It is believed that the cellular immune response is important in the control of the disease and it has recently been demonstrated that in mice a peptide (P10) from the gp43 protein (the main antigen used for immunodiagnose) is capable of triggering a protective cellular immune response.

The dimorphic fungus *P. brasiliensis* is found as a mycelium (M) or as a spore when grown in vitro at room temperature or in its natural habitat. When it is shifted to temperatures close to 37°C it is found as a yeast, the cell form predominantly found in tissues of patients with PCM. Infection occurs mainly when spores or even fragments of mycelium are inhaled. Upon host infection, *P. brasiliensis* undergoes a dimorphic transition from the mycelium form (or spore) to the yeast form probably triggered by the temperature shift. These observations strongly suggest that the dimorphic transition is an important event in the establishment of the infection as it has also been observed in other pathogenic fungi such as *Candida albicans* and *Histoplasma capsulatum*.

Figura 5.6: Página contendo informações sobre o *P. lutzii*



The Pb01 Metabolic Network Project



- ▶ [About *Paracoccidioides brasiliensis*](#)
- ▶ [Data sets](#)
- ▶ [Metabolic pathways](#)
- ▶ [Methodology](#)
- ▶ [Team](#)
- ▶ [Metabolic network picture](#)

Here are some data sets collected in this work:

- [Fasta file with proteins of Pb01](#)
- [List of enzyme codes of Pb01](#)
- [List of enzyme codes of Pb01 not found on kegg on methabolic pathways of *Escherichia coli*](#)
- [Spreadsheet showing swissprot annotation of hypothetical and non-hypothetical proteins](#)
- [Spreadsheet associating proteins with predicted ECs](#)
- [Spreadsheet with reaction data](#)
- [Spreadsheet listing compounds used in reactions](#)
- [Spreadsheet showing, for each compound, the reactions in which the compound occur](#)

Figura 5.7: Página com dados disponibilizados pelo projeto

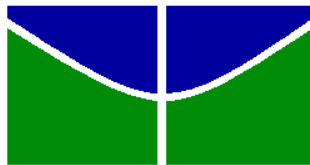
5.2 Discussão

Nesta seção realizamos uma análise preliminar dos resultados obtidos neste trabalho. Analisamos a rede metabólica gerada, com dois focos. Inicialmente, o metabolismo primário é discutido com base em um agrupamento dos principais blocos. Em seguida, identificamos as vias metabólicas do metabolismo secundário.

Idealmente, uma rede metabólica é um modelo fiel e preciso do metabolismo de um organismo. No entanto, as redes metabólicas consistem em aproximações do metabolismo do organismo de estudo. Em geral, uma rede metabólica é incompleta, no sentido de não modelar todas as reações efetuadas em um organismo. Mesmo com essa informação disponível, outros dados, como informações de regulação e fluxo podem ser necessárias. Desta forma, a rede metabólica obtida deve ser aprimorada no sentido de identificar falhas e omissões.

5.2.1 Metabolismo primário

O metabolismo primário consiste nas reações essenciais à vida de qualquer organismo, e portanto é efetuado por todos os organismos vivos. Pode ser agrupado em quatro grandes grupos: síntese/degradação de carboidratos, síntese/degradação de lipídios, síntese/degradação de aminoácidos e síntese/degradação de nucleotídeos. Também está incluso no metabolismo primário o metabolismo energético, responsável pela síntese de adenosina tri-fosfato (ATP), sendo o ATP a principal fonte de energia das células.



The Pb01 Metabolic Network Project



- ▶ [About *Paracoccidioides brasiliensis*](#)
- ▶ [Data sets](#)
- ▶ [Metabolic pathways](#)
- ▶ [Methodology](#)
- ▶ [Team](#)
- ▶ [Metabolic network picture](#)

Metabolic pathways with enzyme codes of Pb01

Here are links for the various metabolic pathways of *Saccharomyces cerevisiae* in which enzyme codes of Pb01 were located. Enzyme codes of *S. cerevisiae* are represented as boxes with a green background, whilst enzyme codes common to both *S. cerevisiae* and Pb01 are represented as green boxes with a red border.

- [sce01100 Metabolic pathways - *Saccharomyces cerevisiae* \(budding yeast\) \(570\)](#)
- [sce01110 Biosynthesis of secondary metabolites - *Saccharomyces cerevisiae* \(budding yeast\) \(237\)](#)
- [sce00190 Oxidative phosphorylation - *Saccharomyces cerevisiae* \(budding yeast\) \(73\)](#)
- [sce00230 Purine metabolism - *Saccharomyces cerevisiae* \(budding yeast\) \(66\)](#)
- [sce00010 Glycolysis / Gluconeogenesis - *Saccharomyces cerevisiae* \(budding yeast\) \(52\)](#)
- [sce00240 Pyrimidine metabolism - *Saccharomyces cerevisiae* \(budding yeast\) \(41\)](#)
- [sce00500 Starch and sucrose metabolism - *Saccharomyces cerevisiae* \(budding yeast\) \(40\)](#)
- [sce00970 Aminoacyl-tRNA biosynthesis - *Saccharomyces cerevisiae* \(budding yeast\) \(37\)](#)
- [sce00020 Citrate cycle \(TCA cycle\) - *Saccharomyces cerevisiae* \(budding yeast\) \(33\)](#)
- [sce00620 Pyruvate metabolism - *Saccharomyces cerevisiae* \(budding yeast\) \(32\)](#)
- [sce00250 Alanine, aspartate and glutamate metabolism - *Saccharomyces cerevisiae* \(budding yeast\) \(31\)](#)
- [sce00330 Arginine and proline metabolism - *Saccharomyces cerevisiae* \(budding yeast\) \(29\)](#)

Figura 5.8: Página com *links* para vias encontradas no KEGG



The Pb01 Metabolic Network Project



- ▶ [About *Paracoccidioides brasiliensis*](#)
- ▶ [Data sets](#)
- ▶ [Metabolic pathways](#)
- ▶ [Methodology](#)
- ▶ [Team](#)
- ▶ [Metabolic network picture](#)

Metabolic pathways with enzyme codes of Pb01

Here are links for the various metabolic pathways of *Saccharomyces cerevisiae* in which enzyme codes of Pb01 were located. The enzymes of Pb01 are represented with a red border.

- [acetate utilization](#)
- [glycerol degradation](#)
- [alanine degradation](#)
- [alanine biosynthesis](#)
- [allantoin degradation](#)
- [arginine degradation \(anaerobic\)](#)
- [spermine and methylthioadenosine biosynthesis](#)
- [arginine biosynthesis](#)
- [chorismate biosynthesis](#)
- [asparagine biosynthesis](#)
- [asparagine degradation](#)
- [aspartate biosynthesis](#)
- [biotin biosynthesis](#)
- [spermidine and methylthioadenosine biosynthesis](#)
- [citrulline biosynthesis](#)
- [cysteine biosynthesis from homocysteine](#)
- [de novo biosynthesis of purine nucleotides](#)
- [removal of superoxide radicals](#)
- [superpathway of ergosterol biosynthesis](#)
- [fatty acid elongation](#)
- [folate biosynthesis](#)
- [glutamine biosynthesis](#)
- [superpathway of glucose fermentation](#)
- [gluconeogenesis](#)
- [lipid-linked oligosaccharide biosynthesis](#)
- [glutamate degradation I](#)
- [glutamate biosynthesis from ammonia](#)
- [glutathione-glutaredoxin redox reactions](#)

Figura 5.9: Página com *links* para vias encontradas no Biocyc

Síntese/degradação de carboidratos

Os carboidratos são a principal fonte de energia para a maioria dos organismos não fotossintéticos. São também utilizados para a fabricação de polímeros insolúveis, sendo esses polímeros são utilizados como elementos estruturais e protetores em diversos organismos. Apresentamos a seguir uma análise das principais vias do metabolismo primário envolvendo carboidratos.

Glicólise

A glicólise é a via metabólica onde ocorre a degradação de açúcares como glicose, frutose, manose e outros para piruvato. A via está completa no *P. lutzii* com exceção da enzima 2.7.9.2, que catalisa a conversão de piruvato para fosfoenolpiruvato. No livro [65], utilizado como referência para estudo das vias, essa reação não é descrita na via da glicólise, também não sabemos dizer se essa reação é efetuada pelo *P. lutzii*.

Gluconeogênese

A gluconeogênese é o processo pelo qual um organismo sintetiza glicose a partir de compostos carbônicos com três ou quatro átomos de carbono para glicose. Esta via está quase completa, com exceção da enzima 2.7.9.2, capaz de converter piruvato em fosfoenolpiruvato. No entanto, essa conversão pode ser feita em dois passos, primeiro transformando piruvato em oxaloacetato, e posteriormente o oxaloacetato em fosfoenolpiruvato, essas reações são catalisadas pelas enzimas 6.4.1.1 e 4.1.1.31, respectivamente. Apenas a enzima 6.4.1.1 foi identificada pelas análises em *P. lutzii*, portanto, a menos que um outro mecanismo de conversão de piruvato em fosfoenolpiruvato seja descoberto, essa via está incompleta pela falta de uma reação.

Ciclo do ácido tricarboxílico

O ciclo do ácido tricarboxílico (TCA) é uma via energética, tendo como entrada grupos acetil, esses grupos são oxidados no TCA e são geradas moléculas de NADH e FADH, essas moléculas são então posteriormente utilizadas na fosforilação oxidativa. Apenas uma reação não está presente no ciclo TCA, mostrada na figura 5.10. Essa reação na verdade é catalisada por um complexo enzimático [65], portanto é possível que não tenha sido encontrada devido ao fato de nenhum complexo proteico ter sido definido. No entanto, as três enzimas do complexo foram identificadas em *P. lutzii* (piruvate dehydrogenase - EC 1.2.4.1, dihydrolipoyl transacetylase - EC 2.3.1.12 e dihydrolipoyl dehydrogenase - EC 1.8.1.4).

Síntese/degradação de glicogênio

A molécula de glicogênio consiste basicamente em polímeros de glicose, e serve para armazenar glicose para posterior utilização. A síntese está completa, pois foi

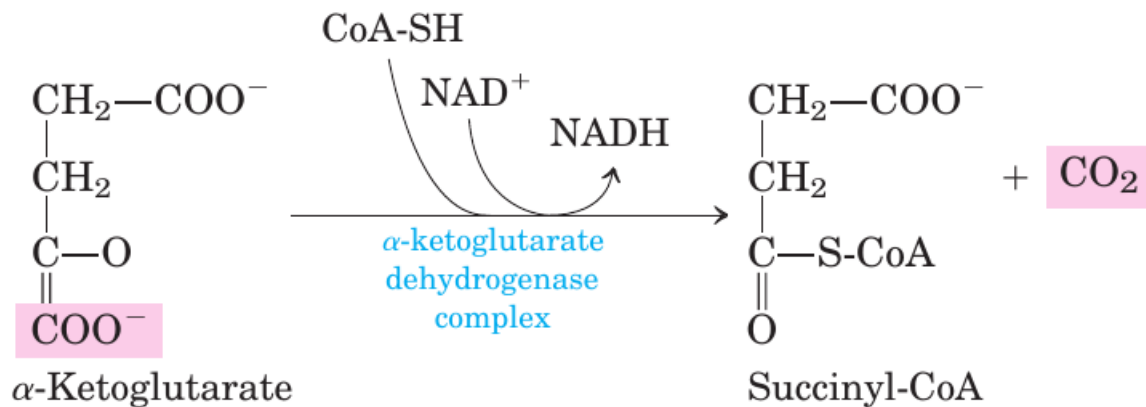


Figura 5.10: Reação não encontrada no ciclo TCA. Figura retirada de [65]

identificada entre as sequências não anotadas de *P. lutzii* a enzima/proteína glicogenina (EC 2.4.1.186). Essa enzima é importante pois é responsável por sintetizar o iniciador a partir de onde as demais enzimas atuarão para formar o glicogênio.

A respeito da degradação de glicogênio, a via está aparentemente completa, com as três principais enzimas do processo identificadas: glicogênio fosforilase, uma enzima para remoção de ramos de glicogênio, e por fim fosfoglicomutase. Aparentemente, a enzima glicogênio fosforilase foi anotada pelo *Pathway Tools* como amido fosforilase, mas isso pode ocorrer devido ao EC 2.4.1.1 possuir mais de um nome, e então acreditamos que a enzima presente seja a glicogênio fosforilase.

Pentoses-fosfato

A via das pentoses-fosfato permite a obtenção de NADPH para combater os efeitos nocivos de radicais livres envolvendo moléculas de oxigênio. Essa via também pode ser utilizada para gerar precursores para a síntese de DNA, RNA e nucleotídeos.

A via das pentoses-fosfato está completa, tanto em respeito ao ramo oxidativo, quanto ao ramo não-oxidativo. Desta forma o *P. lutzii* é capaz de sintetizar NADPH e gerar precursores para a síntese de nucleotídeos.

Síntese/degradação de lipídios

Os lipídios consistem em uma grande diversidade de moléculas, tendo como característica comum a insolubilidade em água. Organismos utilizam lipídios para armazenamento energético, síntese de membranas celulares, cofatores de enzimas, hormônios, isolamento térmico, entre outros. Analisamos então a rede metabólica gerada de *P. lutzii*, focando no metabolismo de lipídios.

A β -oxidação é a via por onde dois carbonos são sucessivamente removidos de uma cadeia lipídica, cada grupo de dois carbonos removidos gera acetil-CoA, que pode ser utilizado no TCA para obter energia. Da família de enzimas acil-CoA desidrogenases, responsáveis por fazer essa remoção, o KAAS anotou a enzima ativa em cadeias de médio comprimento (4 a 14 carbonos) foi identificada. As en-

zimas agindo em cadeias muito longas (12 a 18 carbonos, EC 1.3.8.8) e em cadeias curtas (4 a 8 carbonos, EC 1.3.8.1) foram identificadas em sequências não anotadas de *P. lutzii*. Todas as principais enzimas restantes no processo de β -oxidação foram identificadas. Também foi identificada uma via por onde o *P. lutzii* seria capaz de processar ácidos graxos insaturados, com um número ímpar de carbonos, dependente de di-isomerase. Apenas uma enzima (EC 1.2.1.3) da via da ω -oxidação foi identificada, não sabemos dizer se *P. lutzii* efetivamente utiliza esta via ou se isto é um artefato da anotação.

Para a oxidação de ácidos graxos com um número ímpar de carbonos, três enzimas são necessárias: propionyl-CoA carboxylase (6.4.1.3), methylmalonyl-CoA epimerase (5.1.99.1) e methyl-malonyl-CoA mutase (5.4.99.2). Foi identificada apenas a primeira enzima.

Outras vias detectadas de degradação de lipídios foram: degradação de glicerol, fosfolipases, degradação de triacilglicerol.

Quanto a síntese de ácidos graxos, dos sete ECs que compõem o complexo *fatty acid synthase*, quatro foram identificados no *P. lutzii* (2.3.1.39, 1.1.1.100, 1.3.1.10 e 3.1.2.14). Os ECs 1.3.1.10 e 3.1.2.14 foram identificados por blast [3] com sequências nas quais o KAAS não anotou enzimas. Os ECs não identificados foram: 2.3.1.38, 2.3.1.41 e 4.2.1.59.

Outras vias de síntese/degradação de lipídios identificadas estão na tabela 5.4:

Tabela 5.4: Vias de de lipídios identificadas
Nomes (em inglês)
3-phosphoinositide biosynthesis
biotin-carboxyl carrier protein assembly
cardiolipin biosynthesis II
CDP-diacylglycerol biosynthesis I
choline biosynthesis III
cyclopropane fatty acid (CFA) biosynthesis
dolichol and dolichyl phosphate biosynthesis
ergosterol biosynthesis
fatty acid activation
fatty acids biosynthesis (yeast)
lanosterol biosynthesis
mevalonate pathway I
phosphatidylcholine biosynthesis I
phosphatidylethanolamine biosynthesis I
phosphatidylethanolamine biosynthesis II
triacylglycerol biosynthesis
very long chain fatty acid biosynthesis
zymosterol biosynthesis

Síntese/degradação de aminoácidos

Os aminoácidos são os blocos constituintes das proteínas. Organismos usam as proteínas para as mais diversas funções: estrutural, transporte, defesa. De fato, as próprias enzimas, essenciais para a vida, são proteínas feitas a partir de aminoácidos, portanto os aminoácidos desempenham um papel essencial à manutenção da vida. Aminoácidos podem ser classificados em aminoácidos essenciais e não essenciais, os aminoácidos essenciais são aminoácidos que devem ser adquiridos através da dieta, pois o organismo não é capaz de sintetizá-los, já os aminoácidos não essenciais podem ser sintetizados a partir de moléculas precursoras, e portanto não precisam ser adquiridos por meio da dieta. Embora trabalhos anteriores [8] indiquem que o único aminoácido essencial para *P. lutzii* seja a asparagina, a via de síntese da asparagina está completa na rede metabólica gerada (Figura 5.11).

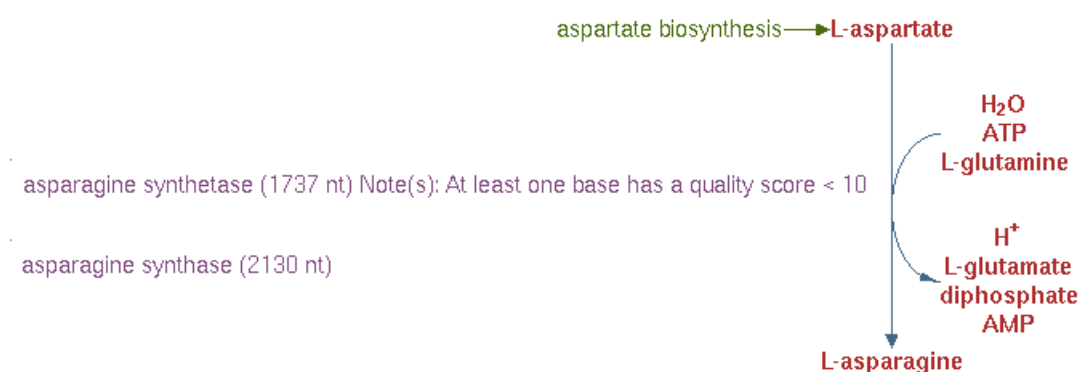


Figura 5.11: Via de síntese da asparagina identificada em *P. lutzii*

A síntese dos aminoácidos se dá principalmente a partir dos aminoácidos glutamato e glutamina. Os outros aminoácidos são criados por diferentes transformações desses dois aminoácidos como base, no entanto, outras vias também existem, portanto é possível haver mais de um meio para sintetizar alguns aminoácidos.

Segundo a rede metabólica obtida, o *P. lutzii* é capaz de sintetizar os seguintes aminoácidos, por pelo menos uma via: alanina, arginina, asparagina, aspartato, glutamato, glicina, histidina, cisteína, isoleucina, leucina, fenilalanina, prolina, serina, treonina e valina. Vias incompletas são: glutamina, triptofano e lisina. Vias não identificadas de síntese de aminoácidos: tirosina e metionina.

Foram encontradas as seguintes vias completas de degradação de aminoácidos em *P. lutzii*: alanina, arginina, asparagina, aspartato, glutamato, glutamina, cisteína, serina, leucina, prolina, treonina, tirosina e valina. Vias incompletas são: metionina (enzima sem EC), triptofano, isoleucina, lisina e fenilalanina. Vias não identificadas: glicina e histidina.

Síntese/degradação de nucleotídeos e nucleosídeos

Vias de síntese completas e quase completas de nucleotídeos e nucleosídeos estão mostradas na tabela 5.5.

Tabela 5.5: Vias completas e quase completas de síntese.

Nomes (em inglês)
5-aminoimidazole ribonucleotide biosynthesis I
adenine and adenosine salvage I
adenine and adenosine salvage III
adenine and adenosine salvage VI
inosine-5'-phosphate biosynthesis II
uridine-5'-phosphate biosynthesis

Vias de degradação completas e quase completas de nucleotídeos e nucleosídeos estão mostradas na tabela 5.6.

Tabela 5.6: Vias completas e quase completas de degradação

Nomes (em inglês)
adenosine nucleotides degradation II
guanosine nucleotides degradation II
purine ribonucleosides degradation to ribose-1-phosphate
pyrimidine ribonucleosides degradation II
S-methyl-5'-thioadenosine degradation II
urate biosynthesis/inosine 5'-phosphate degradation

Fosforilação oxidativa e fermentação

A fosforilação oxidativa é um processo onde elétrons são transferidos entre diferentes carreadores. A cada transferência de elétrons, um certo número de prótons é “bombeado” para fora da membrana interna da mitocôndria. A membrana interna da mitocôndria é impermeável para a passagem de prótons, portanto os prótons “bombeados” não podem retornar pela membrana. Esse processo tem como resultado a formação de um gradiente eletro-químico na membrana interna da mitocôndria. Os prótons podem retornar para o interior da membrana interna através de poros de prótons associados com a enzima ATP sintase, esse movimento de prótons pelos poros fornece a energia necessária para a síntese de ATP.

A fosforilação oxidativa aparenta estar completa no *P. lutzii*, com enzimas para transferência de elétrons de NADH para ubiquinona (ou coenzima Q), de ubiquinol para o citocromo c, e do citocromo c para o oxigênio. Embora o processo envolva mais carreadores de elétrons, acreditamos que a via apresentada está resumida, e que o *P. lutzii* seja capaz de realizar integralmente a fosforilação oxidativa.

Foi identificada uma via de fermentação de piruvato para etanol, condizente com estudos que dizem que o *P. lutzii* efetua metabolismo aeróbico e anaeróbico [32].

Reanotações de enzimas para *P. lutzii*

Na tabela 5.7 são sugeridas anotações para as proteínas do *P. lutzii* anotadas como hipotéticas.

Tabela 5.7: Reanotação para proteínas hipotéticas.

Código (BROAD)	Nomes (em inglês)
PAAG_01870T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_02813T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_02980T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_03233T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_04175T0	linoleate 8R-lipoxygenase EC 1.13.11.60 ou linoleate 10R-lipoxygenase EC 1.13.11.62
PAAG_04275T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10 ou oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_04610T0	pyruvate, water dikinase EC 2.7.9.2
PAAG_07432T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1 ou long-chain acyl-CoA dehydrogenase EC 1.3.8.8
PAAG_09039T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_06838T0	pyruvate, water dikinase EC 2.7.9.2
PAAG_07719T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_00274T0	Cytochrome P450 4A2 EC: 1.14.15.3
PAAG_00927T0	Cytochrome P450 4A25 EC: 1.14.15.3
PAAG_01252T0	Fatty acid oxidation complex subunit alpha EC: 1.1.1.35
PAAG_01265T0	Probable sulfite oxidase, mitochondrial EC: 1.8.3.1
PAAG_02232T0	Cytochrome P450 4A10 EC: 1.14.15.3
PAAG_03398T0	Cytochrome c oxidase subunit 7A EC: 1.9.3.1
PAAG_05405T0	3-hydroxyacyl-CoA dehydrogenase type-2 EC: 1.1.1.35
PAAG_05420T0	Putative endoglucanase X (Fragment) EC: 3.2.1.4
PAAG_06313T0	Cytochrome P450 4A10 EC: 1.14.15.3
PAAG_06816T0	3-hydroxyacyl-CoA dehydrogenase type-2 EC: 1.1.1.35
PAAG_07806T0	Cytochrome P450 4A24 EC: 1.14.15.3
PAAG_08224T0	Phosphatidylglycerophosphatase GEP4, mitochondrial EC: 3.1.3.27
PAAG_08947T0	Fatty acid oxidation complex subunit alpha EC: 1.1.1.35

Na tabela 5.8, são sugeridas anotações para sequências que não tiveram EC atribuído pelo KAAS.

Tabela 5.8: Reanotação para sequências sem ECs

Código (BROAD)	Nomes (em inglês)
PAAG_00121T0	pyruvate, water dikinase EC 2.7.9.2
PAAG_00173T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_00176T0	glycogenin glucosyltransferase EC 2.4.1.186
PAAG_00435T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_01222T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_01516T0	Putative uncharacterized protein <i>Aspergillus niger</i> EC=2.3.1.41
PAAG_01679T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_01790T0	glycogenin glucosyltransferase EC 2.4.1.186
PAAG_01883T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_02512T0	Phenylpyruvate decarboxylase EC 4.1.1.43
PAAG_02691T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_02977T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_03067T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_03490T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_03500T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_04181T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_04787T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_04897T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_05064T0	pyruvate, water dikinase EC 2.7.9.2
PAAG_05359T0	Kynurenine/alpha-aminoadipate aminotransferase mitochondrial, putative) <i>OS=Candida dubliniensis</i> EC 2.6.1.39
PAAG_05376T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10 ou
	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_05416T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_05482T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_05644T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_05782T0	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_05810T0	pyruvate, water dikinase EC 2.7.9.2
PAAG_05989T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_06016T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_06032T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10 ou
	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_06110T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_06135T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_06406T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_06866T0	tryptophan 2,3-dioxygenase EC 1.13.11.11
PAAG_07139T0	pyruvate, water dikinase EC 2.7.9.2
PAAG_07631T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_07665T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_07697T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_07698T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_07721T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_08125T0	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_08320T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10
PAAG_08527T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10 ou
	short-chain acyl-CoA dehydrogenase EC 1.3.8.1
PAAG_08981T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_09046T0	oleoyl-[acyl-carrier-protein] hydrolase EC 3.1.2.14
PAAG_00230T0	Cytochrome P450 4A22 EC: 1.14.15.3
PAAG_00293T0	Fatty acid synthase EC: 2.3.1.41 ou
	enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) EC 1.3.1.10 ou
	Fatty acid synthase EC: 2.3.1.41
PAAG_00629T0	Phthiocerol synthesis polyketide synthase type I PpsC EC: 2.3.1.41
PAAG_01377T0	Glucokinase-1 EC: 2.7.1.2
PAAG_01378T0	Cytochrome P450 4A10 EC: 1.14.15.3
PAAG_02470T0	D-lactate dehydrogenase EC: 1.1.1.28

Código (BROAD)	Nomes (em inglês)
PAAG_02752T0	D-lactate dehydrogenase EC: 1.1.1.28
PAAG_03723T0	Cytochrome P450 4A11 EC: 1.14.15.3
PAAG_03766T0	Probable sulfite oxidase, mitochondrial EC: 1.8.3.1
PAAG_03914T0	Glucokinase EC: 2.7.1.2
PAAG_04963T0	Peroxisomal multifunctional enzyme A EC: 1.1.1.35
PAAG_05218T0	Fatty acid oxidation complex subunit alpha EC: 1.1.1.35
PAAG_05291T0	Fatty acid oxidation complex subunit alpha EC: 1.1.1.35
PAAG_05321T0	Probable sulfite oxidase, mitochondrial EC: 1.8.3.1
PAAG_06064T0	Cytochrome P450 4A14 EC: 1.14.15.3
PAAG_06402T0	Cytochrome P450 4A24 EC: 1.14.15.3
PAAG_07263T0	Cytochrome P450 4A22 EC: 1.14.15.3
PAAG_07727T0	Sulfite oxidase, mitochondrial EC: 1.8.3.1
PAAG_07811T0	Sulfite oxidase EC: 1.8.3.1
PAAG_08057T0	Cytochrome c oxidase polypeptide 5, mitochondrial EC: 1.9.3.1
PAAG_09015T0	Cytochrome P450 4A10 EC: 1.14.15.3

Resumo das vias do metabolismo primário identificadas

- Síntese/degradação de carboidratos
 - Glicólise
 - Gluconeogênese
 - Ciclo do ácido tricarbóxico
 - Síntese/degradação de glicogênio
 - Via das pentoses-fosfato
- Síntese/degradação de lipídios
 - Dos sete ECs que compõem o complexo *fatty acid synthase*, foram identificados quatro (2.3.1.39, 1.1.1.100, 1.3.1.10 e 3.1.2.14), e três não foram identificados (2.3.1.38, 2.3.1.41 e 4.2.1.59).
 - β -oxidação, completa para ácidos graxos saturados com número par de carbonos.
- Síntese/degradação de aminoácidos
 - Na síntese, foram identificadas 15 vias completas, três incompletas (glutamina, triptofano e lisina) e duas não foram identificadas (tirosina e metionina).
 - Na degradação, foram identificadas 13 vias completas, cinco incompletas (metionina, triptofano, isoleucina, lisina e fenilalanina) e duas não foram identificadas (glicina e histidina).
- Síntese/degradação de nucleotídeos e nucleosídeos
 - Na síntese, foram identificadas 6 vias completas e quase completas (5-aminoimidazole ribonucleotide biosynthesis I, adenine and adenosine salvage I, adenine and adenosine salvage III, adenine and adenosine salvage VI, inosine-5'-phosphate biosynthesis II, uridine-5'-phosphate biosynthesis).

- Na degradação foram identificadas 6 vias completas e quase completas (adenosine nucleotides degradation II guanosine nucleotides degradation II purine ribonucleosides degradation to ribose-1-phosphate pyrimidine ribonucleosides degradation II S-methyl-5'-thioadenosine degradation II urate biosynthesis/inosine 5'-phosphate degradation).
- Fosforilação oxidativa: Identificada completamente observando-se que o processo de carreamento de elétrons não foi detalhado.
- Fermentação: Foi identificada a via de oxidação de piruvato para etanol.

5.2.2 Metabolismo secundário

O metabolismo secundário não é essencial para a sobrevivência de um organismo, mas pode contribuir para a adaptação e/ou sobrevivência desse organismo em um ambiente. Como exemplo, a produção de antibióticos por fungos não é necessariamente essencial para a sobrevivência destes, no entanto, aumenta as chances de sobrevivência ao eliminar possíveis organismos concorrentes. A síntese de toxinas também é um exemplo de metabolismo secundário, não sendo essencial para a sobrevivência, mas permitindo a organismos multicelulares se defenderem de ameaças.

A tabela 5.9 mostra as vias identificadas de metabolismo secundário do *P. lutzii*.

Tabela 5.9: Vias do metabolismo secundário.	
Nomes (em inglês)	
	IAA biosynthesis Vit
1D-myo-inositol hexakisphosphate biosynthesis I (from Ins(1,4,5)P3)	
	acetaldehyde biosynthesis II
	antheraxanthin and violaxanthin biosynthesis
	beta-carboline biosynthesis
	betaxanthin biosynthesis
	betaxanthin biosynthesis (via dopamine)
	camptothecin biosynthesis
	D-myo-inositol (1,4,5)-trisphosphate biosynthesis
	D-myo-inositol (1,4,5)-trisphosphate degradation
	epoxysqualene biosynthesis
	mevalonate pathway I
	neurosporaxanthin biosynthesis
	phenylethanol biosynthesis
1D-myo-inositol hexakisphosphate biosynthesis I (from Ins(1,4,5)P3)	
	betanidin degradation
	DIMBOA-glucoside degradation
	mannitol degradation I
	N-acetylglucosamine degradation I
	sorbitol degradation I

5.2.3 Comparação com trabalhos anteriores

Em Felipe et al [32], uma visão geral do metabolismo de *P. brasiliensis* é apresentada. De acordo com esse trabalho, *P. brasiliensis* é capaz de executar glicólise, fosforilação oxidativa e fermentação para álcool, o mesmo tendo sido verificado em nossos resultados. No entanto, tal trabalho apenas fornece 28 ECs diferencialmente expressos entre as duas formas do fungo, já nosso trabalho lista 857 ECs referentes ao metabolismo geral do fungo.

Em Arraes et al [8], um estudo geral do metabolismo de *P. brasiliensis* é feito. No caso do metabolismo de carboidratos, foram identificadas todas as enzimas das vias da glicólise e gluconeogênese, enquanto uma enzima da gluconeogênese não foi identificada em nosso projeto. A síntese de glicogênio foi completamente identificada em Arraes et al [8], enquanto no nosso trabalho, a enzima responsável pela síntese do *primer* iniciador não foi identificada. Tanto a fosforilação oxidativa quanto a síntese de ergosterol foram identificadas em ambos os trabalhos. Referente ao metabolismo de aminoácidos, em nosso trabalho, três aminoácidos têm vias incompletas, e as vias de síntese de dois aminoácidos não foram identificadas, enquanto Arraes et al [8] afirmam que o *P. brasiliensis* é capaz de sintetizar todos os aminoácidos exceto a asparagina, cuja via de síntese foi completamente identificada no nosso trabalho. Por fim, Arraes et al [8] mostram apenas 82 enzimas, enquanto nosso projeto encontrou 857 enzimas.

Existem outros estudos sobre vias metabólicas específicas do gênero *Paracoccidoides* [16, 34, 55, 68], que poderão ser incorporados à rede metabólica do *P. lutzii* em trabalhos posteriores.

Capítulo 6

Conclusão e trabalhos futuros

Neste trabalho, reconstruímos *in silico* uma primeira versão em larga escala da rede metabólica do fungo *Paracoccidioides lutzii* e disponibilizamos uma página com diversas informações sobre o fungo. Particularmente, identificamos vias metabólicas do *P. lutzii* usando os bancos de dados do KEGG dos organismos *Saccharomyces cerevisiae*, *Coccidioides immitis*, *Aspergillus fumigatus*, *Aspergillus nidulans* e *Escherichia coli*, e o banco de dados Biocyc com os organismos *S. cerevisiae* e *E. coli*. Além disso, completamos, pelo menos parcialmente, as vias identificadas no Biocyc, usando sequências do *P. lutzii* não anotadas ou anotadas como proteínas hipotéticas. Listas de enzimas, compostos, reações e compostos/reações foram produzidas para o *P. lutzii*. Por fim, foi produzida uma rede metabólica usando *Pathway Tools*. A página com as informações produzidas nesta dissertação estão em www.biomol.unb.br/Pb01.

Tanto quanto sabemos, não existe uma rede metabólica produzida *in silico* em larga escala. Utilizar FBA poderia ser uma estratégia para completar as vias além do *pathway hole filler* usado neste trabalho. O metabolismo secundário deve ser mais exaustivamente estudado, visando encontrar características específicas do *P. lutzii*. Construir uma visualização espacial da rede para permitir interação com o usuário poderia auxiliar o uso das informações pelos biólogos e permitir aprimorar a rede metabólica do *P. lutzii*. Por fim, experimentos de bancada para confirmar algumas previsões para reações de vias metabólicas de interesse dos biólogos poderiam ser realizados.

Referências

- [454] 454, R. <http://www.454.com>. 2
- [2] Almeida, A. J., Matute, D. R., Carmona, J. A., Martins, M., Torres, I., McEwen, J. G., Restrepo, A., Leão, C., Ludovico, P., and Rodrigues, F. (2007). Genome size and ploidy of *Paracoccidioides brasiliensis* reveals a haploid DNA content: flow cytometry and GP43 sequence analysis. *Fungal genetics and biology*, 44(1):25–31. 26
- [3] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410. 17, 29, 34, 38, 52
- [4] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402. 20
- [5] Andrade, R. V. (2006). *Análise do Transcriptoma e da Expressão diferencial de Genes de Micélio e Levedura de Paracoccidioides brasiliensis*. PhD thesis, Universidade de Brasília - UnB. 2
- [6] Angelo, P. C. S., Nunes-Silva, C. G., Brígido, M. M., Azevedo, J. S. N., Assunção, E. N., Sousa, A. R. B., Patrício, F. J. B., Rego, M. M., Peixoto, J. C. C., Oliveira, W. P., Freitas, D. V., Almeida, E. R. P., Viana, A. M. H. A., Souza, A. F. P. N., Andrade, E. V., Acosta, P. O. A., Batista, J. S., Walter, M. E. M. T., Leomil, L., Anjos, D. A. S., Coimbra, R. C. M., Barbosa, M. H. N., Honda, E., Pereira, S. S., Silva, A., Pereira, J. O., Silva, M. L., Marins, M., Holanda, F. J., Abreu, R. M. M., Pando, S. C., Gonçalves, J. F. C., Carvalho, M. L., Leal-Mesquita, E. R. R. B. P., da Silveira, M. A., Batista, W. C., Atroch, A. L., França, S. C., Porto, J. I. R., Schneider, M. P. C., and Astolfi-Filho, S. (2008). Guarana (*Paullinia cupana* var. *sorbilis*), an anciently consumed stimulant from the Amazon rain forest: the seeded-fruit transcriptome. *Plant cell reports*, 27(1):117–24. 2
- [7] Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L.-S. L. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 32(Database issue):D115–D119. 21, 40

- [8] Arraes, F. B. M., Benoliel, B., Burtet, R. T., Costa, P. L. N., Galdino, A. S., Lima, L. H. A., Marinho-Silva, C., Oliveira-Pereira, L., Pfrimer, P., Procópio-Silva, L., Reis, V. C.-B., and Felipe, M. S. S. (2005). General metabolism of the dimorphic and pathogenic fungus *Paracoccidioides brasiliensis*. *Genetics and Molecular Research GMR*, 4(2):290–308. 4, 31, 32, 53, 59
- [9] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nature Genetics*, 25(1):25–29. 29
- [10] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1):75. 20
- [11] Bagagli, E., Franco, M., Bosco, S. M. G., Hebel-Barbosa, F., Trinca, L. A., and Montenegro, M. R. (2003). High frequency of *paracoccidioides brasiliensis* infection in armadillos (*Dasypus novemcinctus*): an ecological study. *Medical mycology official publication of the International Society for Human and Animal Mycology*, 41(3):217–223. 25
- [12] Bairoch, a. (2000). The ENZYME database in 2000. *Nucleic acids research*, 28(1):304–5. 20
- [13] Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2010). The Pfam protein families database. *Nucleic Acids Research*, 38(Database issue):D211–D222. 17
- [14] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242. 7
- [15] Bernal, A., Ear, U., and Kyrpides, N. (2001). Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Research*, 29(1):126–127. 1
- [16] Biondo, G. A., Dias-Melicio, L. A., Bordon-Graciani, A. P., Kurokawa, C. S., and Soares, A. M. V. C. (2012). Production of leukotriene B4 by *Paracoccidioides brasiliensis*. *Yeast*, 29:201–208. 59
- [17] Blattner, F. R. (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277(5331):1453–1462. 22

- [Broad Institute of Harvard and MIT] Broad Institute of Harvard and MIT. *Paracoccidioides brasiliensis* Sequencing Project. <http://www.broadinstitute.org/>. 4, 32
- [19] Brummer, E., Castaneda, E., and Restrepo, A. (1993). Paracoccidioidomycosis: an update. *Clinical microbiology reviews*, 6(2):89–117. viii, 26, 27, 28
- [20] Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., Walk, T. C., Zhang, P., and Karp, P. D. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 36(Database issue):D623–31. viii, 12, 14
- [21] Claudel-Renard, C., Chevalet, C., Faraut, T., and Kahn, D. (2003). Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Research*, 31(22):6633–6639. 20
- [22] Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921. 1
- [23] Conti-Díaz, I. A. and Rilla, F. D. (1989). Hipótesis sobre el nicho ecológico de *Paracoccidioides brasiliensis*. *Revista Médica del Uruguay*, 5:97–103. 26
- [24] Cottret, L. and Jourdan, F. (2010). Graph methods for the investigation of metabolic networks in parasitology. *Parasitology*, 137(9):1393–1407. 19
- [25] Desjardins, C. A., Champion, M. D., Holder, J. W., Muszewska, A., Goldberg, J., Bailão, A. M., Brigido, M. M., Ferreira, M. E. D. S., Garcia, A. M., Grynberg, M., Gujja, S., Heiman, D. I., Henn, M. R., Kodira, C. D., León-Narváez, H., Longo, L. V. G., Ma, L.-J., Malavazi, I., Matsuo, A. L., Morais, F. V., Pereira, M., Rodríguez-Brito, S., Sakthikumar, S., Salem-Izacc, S. M., Sykes, S. M., Teixeira, M. M., Vallejo, M. C., Walter, M. E. M. T., Yandava, C., Young, S., Zeng, Q., Zucker, J., Felipe, M. S., Goldman, G. H., Haas, B. J., McEwen, J. G., Nino-Vega, G., Puccia, R., San-Blas, G., Soares, C. M. D. A., Birren, B. W., and Cuomo, C. A. (2011). Comparative genomic analysis of human fungal pathogens causing Paracoccidioidomycosis. *PLoS Genetics*, 7(10):e1002345. 27
- [26] Dobson, P. D., Smallbone, K., Jameson, D., Simeonidis, E., Lanthaler, K., Pir, P., Lu, C., Swainston, N., Dunn, W. B., Fisher, P., Hull, D., Brown, M., Oshota, O., Stanford, N. J., Kell, D. B., King, R. D., Oliver, S. G., Stevens, R. D., and Mendes, P. (2010). Further developments towards a genome-scale metabolic model of yeast. *BMC systems biology*, 4(1):145. 17, 23
- [27] Duarte, N. C., Herrgård, M. J., and Palsson, B. O. (2004). Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, 14(7):1298–309. 22
- [28] Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–7. 20

- [29] Ewing, B., Hillier, L., and Wendl, M. C. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, 8(3):175–185. 29
- [30] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., and Ø, B. P. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(121):121. viii, 15, 16, 22
- [31] Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Bernhard (2011). Reconstruction of biochemical networks in microbial organisms. *Nature Reviews Microbiology*, 7(2):129–143. 16, 17, 19
- [32] Felipe, M. S. S., Andrade, R. V., Arraes, F. B. M., Nicola, A. M., Maranhão, A. Q., Torres, F. a. G., Silva-Pereira, I., Poças-Fonseca, M. J., Campos, E. G., Moraes, L. M. P., Andrade, P. a., Tavares, A. H. F. P., Silva, S. S., Kyaw, C. M., Souza, D. P., Pereira, M., Jesuino, R. S. a., Andrade, E. V., Parente, J. a., Oliveira, G. S., Barbosa, M. S., Martins, N. F., Fachin, A. L., Cardoso, R. S., Passos, G. a. S., Almeida, N. F., Walter, M. E. M. T., Soares, C. M. a., Carvalho, M. J. a., and Brígido, M. M. (2005). Transcriptional profiles of the human pathogenic fungus *Paracoccidioides brasiliensis* in mycelium and yeast cells. *The Journal of biological chemistry*, 280(26):24706–14. 2, 3, 29, 32, 54, 59
- [33] Felipe, M. S. S., Andrade, R. V., Petrofeza, S. S., Maranhão, A. Q., Torres, F. A. G., Albuquerque, P., Arraes, F. B. M., Arruda, M., Azevedo, M. O., Baptista, A. J., Bataus, L. A. M., Borges, C. L., Campos, E. G., Cruz, M. R., Daher, B. S., Dantas, A., Ferreira, M. A. S. V., Ghil, G. V., Jesuino, R. S. A., Kyaw, C. M., Leitão, L., Martins, C. R., Moraes, L. M. P., Neves, E. O., Nicola, A. M., Alves, E. S., Parente, J. A., Pereira, M., Poças-Fonseca, M. J., Resende, R., Ribeiro, B. M., Saldanha, R. R., Santos, S. C., Silva-Pereira, I., Silva, M. A. S., Silveira, E., Simões, I. C., Soares, R. B. A., Souza, D. P., De-Souza, M. T., Andrade, E. V., Xavier, M. A. S., Veiga, H. P., Venancio, E. J., Carvalho, M. J. A., Oliveira, A. G., Inoue, M. K., Almeida, N. F., Walter, M. E. M. T., Soares, C. M. A., and Brígido, M. M. (2003). Transcriptome characterization of the dimorphic and pathogenic fungus *Paracoccidioides brasiliensis* by EST analysis. *Yeast Chichester England*, 20(3):263–271. 2, 3, 32
- [34] Fernandes, L., Araújo, M. A. M., Amaral, A., Reis, V. C. B., Martins, N. F., and Felipe, M. S. S. (2005). Cell signaling pathways in *Paracoccidioides brasiliensis* - inferred from comparisons with other fungi. *Genetics and Molecular Research*, 4(2):216–231. 59
- [35] Förster, J., Famili, I., Fu, P., Palsson, B. O., and Nielsen, J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome research*, 13(2):244–53. 16, 19, 22, 32

- [36] Franco, M., Peracoli, M. T., Soares, A., Montenegro, R., Mendes, R. P., and Meira, D. A. (1987). Host-parasite relationships in paracoccidioidomycosis. *Current Topics in Medical Mycology*, 5(1):115–149. 28
- [37] G Junior, D. S., Araújo, F. R., Almeida Junior, N. F., Adi, S. S., Cheung, L. M., Frago, S. P., Ramos, C. A. N., Oliveira, R. H. M. D., Santos, C. S., Bacanelli, G., Soares, C. O., Rosinha, G. M. S., and Fonseca, A. H. (2010). Analysis of membrane protein genes in a brazilian isolate of *Anaplasma marginale*. *Memórias do Instituto Oswaldo Cruz*, 105(7):843–849. 2
- [38] Heavner, B. D., Smallbone, K., Barker, B., Mendes, P., and Walker, L. P. (2012). Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC systems biology*, 6(1):55. 17, 23
- [39] Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H. V., Kırdar, B. I., Penttilä, M., Klipp, E., Palsson, B. O., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J., and Kell, D. B. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology*, 26(10):1155–1160. 17, 22
- [40] Huang, X. and Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. *Genome Research*, 9(9):868–877. 29
- [41] Hucka, M., Finney, A., Sauro, H., Bolouri, H., and Doyle, J. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531. 17
- [Illumina] Illumina. <http://www.illumina.com>. 2
- [IUBMB] IUBMB. International union of biochemistry and molecular biology. <http://www.iubmb.org>. 7
- [44] Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30. viii, 12, 13, 43
- [45] Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I. M., and Caspi, R. (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, 11(1):40–79. 21, 32
- [46] Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Collado-Vides, J., Paley, S. M., Pellegrini-Toole, A., Bonavides, C., and Gama-Castro, S. (2002). The EcoCyc database. *Nucleic Acids Research*, 30(1):56–58. 15

- [47] Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muñiz Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., Gunsalus, R. P., and Karp, P. D. (2011). EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Research*, 39(Database issue):D583–90. viii, 15
- [48] Koonin, E. V. (2002). The clusters of orthologous groups (COGs) database: Phylogenetic classification of proteins from complete genomes. *Antenna*, pages 1–6. 29
- [49] Kuepfer, L., Sauer, U., and Blank, L. M. (2005). Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Research*, 15(10):1421–1430. 22
- [50] Lewin, B. (2003). *Genes VIII*. Benjamin Cummings, united states ed edition. 1
- [51] Lutz, A. (1908). Uma mycose pseudococcidica localizada na boca e observada no Brasil. contribuição ao conhecimento das hyphoblastomycoses americanas. *Brasil Med*, 22:121–141. 25
- [52] Ma, H. and Zeng, A.-P. (2003). Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 19(2):270–277. 22
- [53] Ma, L.-J. and Henn, M. (2006). Comparative genomics of *Coccidioides* and other pathogenic dimorphic fungi. Technical report, Broad Institute. 3
- [54] Majewski, R. A. and Domach, M. M. (1990). Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnology and Bioengineering*, 35(7):732–738. 22
- [55] Marcos, C. M., Silva, J. F., Oliveira, H. C., Silva, R. A. M., Mendes-Giannini, M. J. S., and Fusco-Almeida, A. M. (2012). Surface expressed enolase contributes to the adhesion of *Paracoccidioides brasiliensis* to host cells. *FEMS Yeast Research*, (12):557–570. 59
- [56] Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9(1):387–402. 2
- [57] Marini, M. M., Zanforlin, T., Santos, P. C., Barros, R. R., Guerra, A. C., Puccia, R., Felipe, M. S., Brigido, M., Soares, C. M., Ruiz, J. C., Silveira, J. F., and Cisalpino, P. S. (2010). Identification and characterization of Tc1/mariner-like DNA transposons in genomes of the pathogenic fungi of the *Paracoccidioides* species complex. *BMC Genomics*, 11(1):130. 27
- [58] Martinez, R. (2010). Paracoccidioidomycosis: the dimension of the problem of a neglected disease. *Revista da Sociedade Brasileira de Medicina Tropical*, 43(4):480. 28

- [59] Matute, D. R., McEwen, J. G., Puccia, R., Montes, B. A., San-Blas, G., Bagagli, E., Rauscher, J. T., Restrepo, A., Morais, F., Niño Vega, G., and Taylor, J. W. (2006). Cryptic speciation and recombination in the fungus *Paracoccidioides brasiliensis* as revealed by gene genealogies. *Molecular Biology and Evolution*, 23(1):65–73. 27
- [60] McEwen, J. G., Bedoya, V., Patiño, M. M., Salazar, M. M., and Restrepo, A. M. (1987). Experimental murine paracoccidioidomycosis induced by the inalation of conidia. *Medical Mycology*, 25(3):165–175. 27
- [61] Medoff, G., Painter, A., and Kobayashi, G. S. (1987). Mycelial to yeast-phase transitions of the dimorphic fungi *Blastomyces dermatitidis* and *Paracoccidioides brasiliensis*. *Journal Of Bacteriology*, 169(9):4055–4060. 28
- [62] Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A., and Shlomi, T. (2011). Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proceedings of the National Academy of Sciences*, 109(1):1–6. 22
- [63] Montoya, A. E., Moreno, M. N., Restrepo, A., and McEwen, J. G. (1997). Electrophoretic karyotype of clinical isolates of *Paracoccidioides brasiliensis*. *Fungal genetics and biology FG B*, 21(2):223–227. 26
- [64] Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35(Web Server issue):W182–W185. 12, 32, 38
- [65] Nelson, D. L. and Cox, M. C. (2005). *Principles of biochemistry*. W. H. Freeman, 4 edition. viii, 1, 6, 50, 51
- [66] of Biochemistry, I. U. and Biology., M. (1992). *Enzyme nomenclature. Recommendations 1992*. Academic Press, 1 edition. 7
- [67] Orth, J. D., Conrad, T. M., Na, J., Lerman, J. a., Nam, H., Feist, A. M., and Palsson, B. O. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Molecular systems biology*, 7(535):535. 22
- [68] Parente, A. F. A., Bailão, A. M., Borges, C. L., Parente, J. A., Magalhães, A. D., Ricart, C. A. O., and Soares, C. M. A. (2011). Proteomic analysis reveals that iron availability alters the metabolic status of the pathogenic fungus *Paracoccidioides brasiliensis*. *PloS one*, 6(7):e22810. 59
- [69] Pinney, J. W., Shirley, M. W., McConkey, G. A., and Westhead, D. R. (2005). metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Research*, 33(4):1399–1409. 20
- [70] Pramanik, J. and Keasling, J. D. (1998). Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnology and Bioengineering*, 60(2):230–8. 22

- [Projeto Biofoco] Projeto Biofoco. <http://www.biofoco.org/biofoco3>. 2
- [Projeto Genoma Anaplasma] Projeto Genoma Anaplasma. <http://www.biomol.unb.br/anaplasma>. 2
- [Projeto Genoma Guaraná] Projeto Genoma Guaraná. <http://www.biomol.unb.br/GR>. 2
- [Projeto Genoma Pb] Projeto Genoma Pb. <http://www.biomol.unb.br/Pb>. 3
- [75] Restrepo, A., McEwen, J. G., and Castañeda, E. (2001). The habitat of *paracoccidioides brasiliensis*: how far from solving the riddle? *Medical mycology official publication of the International Society for Human and Animal Mycology*, 39(3):233–241. 26, 27
- [76] Ricci, G., Mota, F. T., Wakamatsu, A., Serafim, R. C., Borra, R. C., and Franco, M. (2004). Canine paracoccidioidomycosis. *Medical mycology official publication of the International Society for Human and Animal Mycology*, 42(4):379–383. 25
- [77] Salazar, M. E. and Restrepo, A. (1985). Morphogenesis of the mycelium-to-yeast transformation in *Paracoccidioides brasiliensis*. *Sabouraudia Journal Of Medical And Veterinary Mycology*, 23(1):7–11. 25
- [78] Salazar, M. E., Restrepo, A., and Stevens, D. A. (1988). Inhibition by estrogens of conidium-to-yeast conversion in the fungus *Paracoccidioides brasiliensis*. *Infection and Immunity*, 56(3):711–713. 28
- [79] Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5463–5467. 2
- [80] Schomburg, I., Chang, A., and Schomburg, D. (2002). BRENDA, enzyme data and metabolic information. *Nucleic Acids Research*, 30(1):47–49. 40
- [81] Setubal, J. and Meidanis, J. (1997). *Introduction to Computational Molecular Biology*. Number 3. PWS Publishing Company. 2
- [82] Sharpton, T. J., Stajich, J. E., Rounsley, S. D., Gardner, M. J., Wortman, J. R., Jordan, V. S., Maiti, R., Kodira, C. D., Neafsey, D. E., Zeng, Q., Hung, C.-Y., McMahan, C., Muszewska, A., Grynberg, M., Mandel, M. A., Kellner, E. M., Barker, B. M., Galgiani, J. N., Orbach, M. J., Kirkland, T. N., Cole, G. T., Henn, M. R., Birren, B. W., and Taylor, J. W. (2009). Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Research*, 19(10):1722–1731. 25
- [83] Simpson, A. J., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., Alves, L. M., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bové, J. M., Briones, M. R., Bueno, M. R., Camargo, A. A., Camargo, L. E., Carraro, D. M., Carrer, H., Colauto, N. B., Colombo, C., Costa, F. F., Costa, M. C., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M.,

Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A. P., Ferreira, A. J., Ferreira, V. C., Ferro, J. A., Fraga, J. S., França, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J. P., Krieger, J. E., Kuramae, E. E., Laigret, F., Lambais, M. R., Leite, L. C., Lemos, E. G., Lemos, M. V., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, A. M., Madeira, H. M., Marino, C. L., Marques, M. V., Martins, E. A., Martins, E. M., Matsukuma, A. Y., Menck, C. F., Miracca, E. C., Miyaki, C. Y., Monteriro-Vitorello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L., Netto, L. E., Nhani, A., Nobrega, F. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A., Pereira, H. A., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M Rosa, A. J., de Rosa, V. E., de Sá, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C., da Silva, A. M., da Silva, F. R., da Silva, W. A., da Silveira, J. F., Silvestri, M. L., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tshako, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J., and Setubal, J. C. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature*, 406(6792):151–9. 1

[SOLiD] SOLiD, A. <http://www.appliedbiosystems.com.br/site/abhome.jsp>. 2

[85] Sundquist, A., Ronaghi, M., Tang, H., Pevzner, P., and Batzoglou, S. (2007). Whole-Genome Sequencing and Assembly with High-Throughput, Short-Read Technologies. *PLoS ONE*, 2(5):14. 2

[86] Teixeira, M. M. (2008). Tipagem molecular e evolução do gênero *Paracoccidioides*. Master's thesis, Universidade de Brasília - UnB. 2, 4

[87] Teixeira, M. M. (2012). *Especiação, genômica e reprodução sexuada no complexo de espécies do gênero Paracoccidioides*. PhD thesis, Programa de Pós-Graduação em Biologia Molecular da Universidade de Brasília. 4

[88] Teixeira, M. M., Theodoro, R. C., De Carvalho, M. J. A., Fernandes, L., Paes, H. C., Hahn, R. C., Mendoza, L., Bagagli, E., San-Blas, G., and Felipe, M. S. S. (2009). Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides* genus. *Molecular Phylogenetics and Evolution*, 52(2):273–283. 27

[89] Theodoro, R. C., Bagagli, E., and Oliveira, C. (2008). Phylogenetic analysis of PRP8 intein in *Paracoccidioides brasiliensis* species complex. *Fungal genetics and biology FG B*, 45(9):1284–1291. 27

[90] Theodoro, R. C., Teixeira, M. D. M., Felipe, M. S. S., Paduan, K. D. S., Ribolla, P. M., San-Blas, G., and Bagagli, E. (2012). Genus *paracoccidioides*: species recognition and biogeographic aspects. *PloS one*, 7(5):e37694. 27

[91] Varma, A., Boesch, B. W., and Palsson, B. O. (1993). Biochemical production capabilities of *Escherichia coli*. *Biotechnology and Bioengineering*, 42(1):59–73. 22

[92] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. a., Holt, R. a., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, a. G., Nadeau, J., McKusick, V. a., Zinder, N., Levine, a. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, a., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, a., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, a. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. a., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, a. K., Narayan, V. a., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, a., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, a., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, a., Woodage, T., Ali, F., An, H., Awe, a., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Besson, K., Busam, D., Carver, a., Center, a., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, a., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, a., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, a., Mi, H., Lazareva, B., Hatton, T., Narechania, a., Diemer, K., Muruganujan, a., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, a., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, a., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, a., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, a., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, a., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, a., Zandieh, a., and Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507):1304–51. 1

[93] Watson, J. D. and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738. 1

- [94] Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., Chu, A. M., Connelly, C., Davis, K., Dietrich, F., Dow, S. W., El Bakkoury, M., Foury, F., Friend, S. H., Gentalen, E., Giaever, G., Hegemann, J. H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D. J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J. L., Riles, L., Roberts, C. J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R. K., Véronneau, S., Voet, M., Volckaert, G., Ward, T. R., Wysocki, R., Yen, G. S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M., and Davis, R. W. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, 285(5429):901–906. 16