

MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

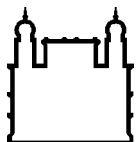
Doutorado em Biologia Computacional e Sistemas

BUSCA DE ALVOS PROTEICOS PARA DESENVOLVIMENTO DE
INIBIDORES ENZIMÁTICOS EM SISTEMAS HOSPEDEIRO-
PARASITO

LARISSA CATHARINA COSTA

Rio de Janeiro

Dezembro de 2017



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

Larissa Catharina Costa

Busca de alvos proteicos para desenvolvimento de inibidores enzimáticos em sistemas hospedeiro-parasito

Tese ou dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutor em Biologia Computacional e Sistemas.

Orientador (es): Prof. Dr. Nicolas Carels
Prof. Dr. Marcio Argollo de Menezes

RIO DE JANEIRO

Dezembro de 2017

Catharina Costa, Larissa.

Busca de alvos proteicos para desenvolvimento de inibidores enzimáticos em sistemas hospedeiro-parasito. / Larissa Catharina Costa. - Rio de Janeiro, 2017.

254 f.; il.

Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2017.

Orientador: Nicolas Carels.

Co-orientador: Marcio Argollo de Menezes.

Bibliografia: f. 141-200

1. Alvos proteicos. 2. Inibidores enzimáticos. 3. Hospedeiro-parasito. 4. Enzimas específicas e análogas. 5. Análise de Balanço de Fluxo - FBA. I. Título.

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: LARISSA CATHARINA COSTA

**BUSCA DE ALVOS PROTEICOS PARA DESENVOLVIMENTO DE INIBIDORES EM
SISTEMAS HOSPEDEIRO-PARASITO**

ORIENTADOR (ES): Prof. Dr. Nicolas Carels

Prof. Dr. Marcio Argollo de Menezes

Aprovada em: ____/____/____

EXAMINADORES:

Prof. Dr. Marcio Lourenço Rodrigues – Presidente (Centro de Desenvolvimento Tecnológico em Saúde – CDTS. Fundação Oswaldo Cruz – Fiocruz.)

Prof. Dr. Eduardo Caio Torres dos Santos (Instituto Oswaldo Cruz – IOC. Fundação Oswaldo Cruz – Fiocruz.) *Membro titular*

Prof. Dr. Fabricio Alves Barbosa da Silva (Programa de Computação Científica – PROCC. Fundação Oswaldo Cruz – Fiocruz.) *Membro titular*

Prof. Dr. Antonio Basílio de Miranda (Instituto Oswaldo Cruz – IOC. Fundação Oswaldo Cruz – Fiocruz.) *Suplente*

Prof. Dr. Daniel Savignon Marinho (Centro de Desenvolvimento Tecnológico em Saúde – CDTS. Fundação Oswaldo Cruz – Fiocruz.) *Suplente*

Rio de Janeiro, 15 de Dezembro de 2017

Aos meus Pais e ao meu Príncipe.

AGRADECIMENTOS

Durante a minha formação diversas pessoas contribuíram para o meu crescimento seja este pessoal ou profissional. Portanto, gostaria imensamente de agradecer...

Ao meu querido Deus que está sempre cuidando de mim; está me protegendo, guiando meus passos e fazendo morada em meu coração.

Aos meus pais, Eliana e Juscelino, pelo amor, orações e apoio intenso. Apesar das dificuldades, conseguimos mais uma vitória! Eu amo vocês!

Ao meu Príncipe Rafael por todo amor e cumplicidade; por estar sempre ao meu lado me incentivando e ajudando a tornar os meus dias mais felizes. Eu te amo!

À minha família por todo o apoio, em especial à minha tia Nelide (Dida) e à minha estrelinha (Estela) que iluminam a minha vida.

À minha avó Nadir que está sempre comigo. Seus ensinamentos permanecem comigo.

A todos os meus amigos pela amizade sincera. “A melhor parte da vida de uma pessoa está nas suas amizades.” Abraham Lincoln

Aos meus orientadores Dr. Nicolas Carels e Dr. Marcio Argollo por todos os conhecimentos e ensinamentos gerados, disponibilidade, empenho dedicado à elaboração deste trabalho. Obrigada por tudo, pois sem vocês nada disso seria possível.

À Alessandra Jordano e Tatiana Tilli pela amizade, carinho e principalmente pelo zelo que sempre tiveram por mim. De companheiras de trabalho vocês se tornaram minhas amigas especiais. Sentirei muita saudade.

Aos amigos do Centro de Desenvolvimento Tecnológico em Saúde (CDTS) que me receberam de braços abertos. Fui muito bem recebida por todos do CDTS. Guardo todos no coração. Obrigada pela acolhida.

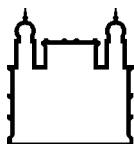
À amiga e pesquisadora Ana Carolina Ramos Guimarães por acreditar em mim e me apresentar o incrível mundo da Bioinformática. Obrigada, Carol, por tudo.

Ao Dr. Carlyle Ribeiro pelas contribuições nos experimentos de modelagem 3D e análises físico-químicas.

Aos membros banca de avaliação, Dr. Marcio Lourenço, Dr. Eduardo Caio, Dr. Fabricio Alves, Dr. Antonio Basílio e Dr. Daniel Marinho, pela participação. A contribuição de cada um é primordial e agregará um imenso valor à tese.

Aos membros da Pós-graduação em Biologia Computacional e Sistemas em especial à Rose Pani, pois sem seu auxílio eu ficaria completamente perdida. Você é essencial!

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES e Instituto National de ciência e Technologia de Inovação em doenças Negligenciadas (INCT-IDN) pelo auxílio financeiro.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

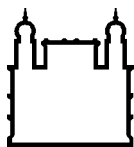
BUSCA DE ALVOS PROTEICOS PARA DESENVOLVIMENTO DE INIBIDORES ENZIMÁTICOS EM SISTEMAS HOSPEDEIRO-PARASITO

RESUMO

TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Larissa Catharina Costa

Diante da necessidade de uma metodologia para buscar e avaliar a relevância do desenvolvimento de inibidores para enzimas específicas em um organismo patógeno que pudesse ser aplicada em sistemas hospedeiro-parasito em geral foram utilizados dois modelos; um envolvendo um vertebrado, *Homo sapiens*, e um protozoário, *Leishmania major*, e o outro envolvendo cinco espécies de plantas, *Arabidopsis thaliana*, *Brassica rapa*, *Glycine max*, *Jatropha curcas* e *Ricinus communis*, e o fungo *Fusarium oxysporum*. Através de aplicações de técnicas tradicionais de comparação de homologia de sequência por busca de similaridade e modelagem de Markov, a metodologia caracteriza o tipo de especificidade enzimática associado aos alvos proteicos que podem ser considerados para o controle da leishmaniose e da podridão radicular causadas por *L. major* e por *F. oxysporum*, respectivamente. Além disso, para predizer computacionalmente os genes críticos através de simulações baseadas em restrições da produção de biomassa utilizando a reconstrução metabólica de *F. oxysporum* foi utilizada *Flux Balance Analysis*, seguido de modelagem tridimensional dos genes críticos no caso de *F. oxysporum*. A metodologia de busca por enzimas específicas permitiu de identificar 40 enzimas estritamente específicas de *L. major* no seu sistema parasitário com *H. sapiens*, dentre essas é podemos sugerir as enzimas esterol 24-C-metiltransferase, piruvato fosfato diquinase, tripanotona sintetase e RNA-edição ligase como alvos apropriados para o desenvolvimento de fármacos para leishmaniose. Ao aplicar a mesma metodologia no modelo planta-fungo, foi possível identificar 30 enzimas específicas para *F. oxysporum* em relação às cinco espécies de plantas estudadas. A modelagem por FBA permitiu identificar 39 enzimas de *F. oxysporum*, críticas para a produção de biomassa, cuja inibição teria potencialidade de desarticular a rede metabólica do fungo *in vivo*, sendo que duas, F9F4G5 e F9FSB6 (consideradas como sendo análogas com suas contrapartes em *A. thaliana*) seriam apropriada para o desenvolvimento de inibidores específico do fungo. Por último, somente foi possível modelar a estrutura 3D do par de uma das enzimas consideradas críticas para a produção de biomassa pelo fungo F9F4G5 (EC: 3.5.4.26), de *F. oxysporum* com seu respectivo par Q8GWP5 (EC: 3.5.4.26) em *A. thaliana*. Portanto, a enzima de *F. oxysporum*, F9F4G5, pode ser considerada um alvo apropriado à inibição uma vez que está envolvida no crescimento do fungo e regulação de processos biológicos essenciais.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

SEARCH FOR PROTEIN TARGETS FOR DEVELOPMENT OF ENZYMATIC INHIBITORS IN HOST-PARASITE SYSTEMS

ABSTRACT

PHD THESIS IN COMPUTATIONAL AND SYSTEMS BIOLOGY

Larissa Catharina Costa

In view of the need for a methodology to search for and evaluate the relevance of inhibitor development for specific enzymes of a pathogenic organism that could be applied in host-parasite systems, two models were used; one involving a vertebrate, *Homo sapiens*, and a protozoa, *Leishmania major*, and the other involving five species of host plants, *Arabidopsis thaliana*, *Brassica rapa*, *Glycine max*, *Jatropha curcas* and *Ricinus communis*, and the fungus *Fusarium oxysporum*. Through the application of traditional techniques of sequence homology comparison by similarity search and Markov modeling, the methodology characterizes the type of enzymatic specificity of protein targets that can be considered for the control of *Leishmaniasis* and root rot caused by *L. major* and *F. oxysporum*, respectively. In addition, to computationally predict critical genes through simulations based on biomass production constraints using the metabolic reconstruction of *F. oxysporum*, Flux Balance Analysis and three-dimensional modeling of critical fungal proteins were used in sequence. The methodology of specific enzyme characterization allowed us to identify 40 strictly specific enzymes of *L. major* in its parasitic association with *H. sapiens*, among which sterol 24-C-methyltransferase, pyruvate phosphate dikinase, trypanothione synthase and RNA ligase could be considered as suitable targets for the drug development for *Leishmaniasis*. When applying the same methodology in the plant-fungus model, it was possible to identify 30 enzymes specific for *F. oxysporum* in relation to the five species of plant hosts studied. FBA modeling allowed to identify 39 enzymes of *F. oxysporum*, that are critical for biomass production, i.e., their inhibition would have the potential to disarticulate the metabolic network of the fungus *in vivo*. Among these 39 genes, two, F9F4G5 and F9FSB6, were considered analogous to their *A. thaliana* counterparts. Finally, it was possible to model the 3D structure of only one pair (F9F4G5 vs. Q8GWP5) of the analogous enzymes considered critical for the production of biomass by *F. oxysporum*. Therefore, the enzyme, F9F4G5, can be considered a suitable target for *F. oxysporum* inhibition since it is involved in fungal growth as well as in the regulation of essential biological processes.

SUMÁRIO

AGRADECIMENTOS -----	v
RESUMO -----	viii
ABSTRACT -----	ix
SUMÁRIO -----	x
ÍNDICE DE FIGURAS -----	xiv
LISTA DE TABELAS -----	xvi
PREÂMBULO -----	xviii
1. INTRODUÇÃO -----	1
1.1 Alvos em Sistemas Hospedeiro-Parasitos.....	1
1.2 Bioinformática.....	6
1.2.1 Genômica Funcional.....	7
1.3 Seleção do Modelo Hospedeiro – Parasito.....	8
1.3.1 Modelo <i>H. sapiens</i> – <i>L. major</i>	9
1.3.2 <i>H. sapiens</i> – <i>Leishmania</i> : interação hospedeiro – parasito.....	13
1.4 Seleção Modelo Planta – Fungo.....	16
1.4.1 Modelo <i>A. thaliana</i> – <i>F. oxysporum</i>	18
1.4.2 <i>A. thaliana</i> – <i>F. oxysporum</i> : interação hospedeiro – patógeno.....	28
1.5 Modelagem por Análise de Balanço de Fluxo - FBA.....	33
2. OBJETIVOS -----	37
2.1 Objetivo Geral.....	37
2.2 Objetivos Específicos.....	37
3. MATERIAL E MÉTODOS -----	38
3.1 Busca de Enzimas Específicas em <i>L. major</i> (Artigo publicado - Anexo 1).....	38
3.1.1 Conhecimento utilizado.....	38
3.1.2 Processo de depuração de enzimas estritamente específicas.....	39
3.1.3 Mapeamento de proteínas homólogas de <i>L. major</i> no genoma humano.....	43
3.1.4 Consistência em pares homólogos de <i>L. major</i> vs. <i>H. sapiens</i> usando BLAST.....	45
3.1.5 <i>Hits</i> genômicos sem contrapartes em proteínas do Ensembl.....	49
3.1.6 O caso particular da especificidade funcional.....	50
3.1.7 Mapeamento de domínio de proteínas e estrutura secundária.....	51

3.1.8 Detecção de homologia remota com modelos ocultos de Markov.....	51
3.1.9 Significância metabólica.....	52
3.1.10 Comparação de sequências de enzimas estritamente específicas de <i>L. major</i> com outros parasitos.....	53
3.1.11 Análise Filogenética.....	56
3.2 Busca de Enzimas Específicas em <i>F. oxysporum</i> (Confirmação de Submissão de Artigo – Anexo 2).....	57
3.2.1 Conhecimento utilizado.....	57
3.2.2 Pesquisa de enzimas ortólogas em <i>F. oxysporum</i>	58
3.2.3 <i>Hidden Markov model</i>	58
3.2.4 Depuração das enzimas específicas para casos falsos positivos.....	59
3.2.5 Significância metabólica.....	64
3.3 Validação das Enzimas Específicas por Análise de Balanço de Fluxo – FBA.....	64
3.3.1 Análise de balanço de fluxo da reconstrução metabólica de <i>F. oxysporum</i>	64
3.3.2 Detecção de homologia entre genes de <i>F. oxysporum</i> e <i>A. thaliana</i>	66
3.3.3 Modelagem tridimensional das duas enzimas análogas consideradas críticas.....	66
3.3.4 Análises físico-químicas.....	67
4. RESULTADOS	67
4.1 Busca de Enzimas Específicas em <i>L. major</i>	67
4.1.1 Metodologia de filtragem de enzimas estritamente específicas de <i>L. major</i> em comparação com <i>H. sapiens</i>	67
4.1.2 Enzimas estritamente específicas de <i>L. major</i>	77
4.1.3 Relacionamento evolutivo de enzimas estritamente específicas em <i>L. major</i> com outros parasitos humanos.....	86
4.2 Busca de Enzimas Específicas em <i>F. oxysporum</i>	87
4.2.1 Comparação de EC e recuperação da sequência de proteínas.....	87
4.2.2 Busca por enzimas específicas em <i>F. oxysporum</i>	88
4.2.3 Depuração de anotação de homólogos de enzimas hospedeiras.....	90
4.2.4 Significância Metabólica.....	91
4.3 Validação das Enzimas Específicas por Análise de Balanço de Fluxo – FBA.....	94
4.3.1 Análise de balanço de fluxo da reconstrução metabólica de <i>F. oxysporum</i>	94
4.3.2 Detecção de homologia entre genes de <i>F. oxysporum</i> e <i>A. thaliana</i>	97
4.3.3 Modelagem tridimensional das duas enzimas análogas consideradas críticas.....	98
5. DISCUSSÃO	104
5.1 Busca de Enzimas Específicas em <i>L. major</i>	104
5.1.1 Desafios de classificação e ambiguidade de anotação.....	104

5.1.2	Essencialidade enzimática.....	107
5.1.3	Especificidade funcional.....	107
5.1.4	Especificidade estrita.....	108
5.1.5	Redundância genética e resistência parasitária aos fármacos.....	112
5.1.6	Fármacos para enzimas de <i>L. major</i> estritamente específicas.....	113
5.1.7	Distribuição de enzimas estritamente específicas de <i>L. major</i> entre parasitos humanos.....	115
5.1.8	Padrão de expressão das principais enzimas específicas de <i>L. major</i>	116
5.2	Busca de Enzimas Específicas em <i>F. oxysporum</i>	116
5.2.1	Enzimas Principais.....	118
5.2.2	Enzimas Auxiliares.....	123
5.3	Validação das Enzimas Específicas por Análise de balanço de fluxo – FBA.....	126
5.3.1	Análise de balanço de fluxo da reconstrução metabólica de <i>F. oxysporum</i>	126
5.3.2	Genes e atividades de enzimas inseridas na Reconstrução metabólica de <i>F. oxysporum</i>	130
5.3.3	Modelagem tridimensional das duas enzimas análogas consideradas críticas.....	133
6.	PERSPECTIVAS -----	136
7.	CONCLUSÃO -----	138
8.	REFERÊNCIAS BIBLIOGRÁFICAS -----	141
9.	ANEXO 1 -----	201
10.	ANEXO 2 -----	223
11.	ANEXO 3	232
1.	INTRODUÇÃO	1
1.1	Alvos em Sistemas Hospedeiro-Parasitos-----	1
1.2	Bioinformática-----	6
1.2.1	Genômica Funcional-----	7
1.3	Seleção do Modelo Hospedeiro – Parasito-----	8
1.3.1	Modelo <i>H. sapiens</i> – <i>L. major</i> -----	9
1.3.1.1	<i>H. sapiens</i> -----	9
1.3.1.2	<i>L. major</i> -----	10
1.3.2	<i>H. sapiens</i> – <i>Leishmania</i> : interação hospedeiro – parasito-----	13
1.4	Seleção Modelo Planta – Fungo-----	16
1.4.1	Modelo <i>A. thaliana</i> – <i>F. oxysporum</i> -----	18
1.4.1.1	<i>A. thaliana</i> -----	18
1.4.1.2	<i>F. oxysporum</i> -----	22

1.4.2	<i>A. thaliana</i> – <i>F. oxysporum</i> : interação hospedeiro – patógeno-----	28
1.5	Modelagem por Análise de Balanço de Fluxo - FBA-----	33
2.	OBJETIVOS	37
2.1	Objetivo Geral-----	37
2.2	Objetivos Específicos-----	37
3.	MATERIAL E MÉTODOS	38
3.1	Busca de Enzimas Específicas em <i>L. major</i> (Artigo publicado - Anexo 1)-----	38
3.1.1	Conhecimento utilizado-----	38
3.1.2	Processo de depuração de enzimas estritamente específicas-----	39
3.1.3	Mapeamento de proteínas homólogas de <i>L. major</i> no genoma humano-----	43
3.1.4	Consistência em pares homólogos de <i>L. major</i> vs. <i>H. sapiens</i> usando BLAST-----	45
3.1.5	<i>Hits</i> genômicos sem contrapartes em proteínas do Ensembl-----	49
3.1.6	O caso particular da especificidade funcional-----	50
3.1.7	Mapeamento de domínio de proteínas e estrutura secundária-----	51
3.1.8	Detecção de homologia remota com modelos ocultos de Markov-----	51
3.1.9	Significância metabólica-----	52
3.1.10	Comparação de sequências de enzimas estritamente específicas de <i>L. major</i> com outros parasitos-----	53
3.1.11	Análise Filogenética-----	56
3.2	Busca de Enzimas Específicas em <i>F. oxysporum</i> (Confirmação de Submissão de Artigo – Anexo 2)-----	57
3.2.1	Conhecimento utilizado-----	57
3.2.2	Pesquisa de enzimas ortólogas em <i>F. oxysporum</i> -----	58
3.2.3	<i>Hidden Markov model</i> -----	58
3.2.4	Depuração das enzimas específicas para casos falsos positivos-----	59
3.2.5	Significância metabólica-----	64
3.3	Validação das Enzimas Específicas por Análise de Balanço de Fluxo – FBA-----	64
3.3.1	Análise de balanço de fluxo da reconstrução metabólica de <i>F. oxysporum</i> -----	64
3.3.2	Detecção de homologia entre genes de <i>F. oxysporum</i> e <i>A. thaliana</i> -----	66
3.3.3	Modelagem tridimensional das duas enzimas análogas consideradas críticas-----	66
3.3.4	Análises físico-químicas-----	67
4.	RESULTADOS	67
4.1	Busca de Enzimas Específicas em <i>L. major</i> -----	67
4.1.1	Metodologia de filtragem de enzimas estritamente específicas de <i>L. major</i> em comparação com <i>H. sapiens</i> -----	67
4.1.2	Enzimas estritamente específicas de <i>L. major</i> -----	77

4.1.3	Relacionamento evolutivo de enzimas estritamente específicas em <i>L. major</i> com outros parasitos humanos-----	86
4.2	Busca de Enzimas Específicas em <i>F. oxysporum</i> -----	87
4.2.1	Comparação de EC e recuperação da sequência de proteínas-----	87
4.2.2	Busca por enzimas específicas em <i>F. oxysporum</i> -----	88
4.2.3	Depuração de anotação de homólogos de enzimas hospedeiras-----	90
4.2.4	Significância Metabólica-----	91
4.3	Validação das Enzimas Específicas por Análise de Balanço de Fluxo – FBA-----	94
4.3.1	Análise de balanço de fluxo da reconstrução metabólica de <i>F. oxysporum</i> -----	94
4.3.2	Detecção de homologia entre genes de <i>F. oxysporum</i> e <i>A. thaliana</i> -----	97
4.3.3	Modelagem tridimensional das duas enzimas análogas consideradas críticas-----	98
5.	DISCUSSÃO	104
5.1	Busca de Enzimas Específicas em <i>L. major</i> -----	104
5.1.1	Desafios de classificação e ambiguidade de anotação-----	104
5.1.2	Essencialidade enzimática-----	107
5.1.3	Especificidade funcional-----	107
5.1.4	Especificidade estrita-----	108
5.1.5	Redundância genética e resistência parasitária aos fármacos-----	112
5.1.6	Fármacos para enzimas de <i>L. major</i> estritamente específicas-----	113
5.1.7	Distribuição de enzimas estritamente específicas de <i>L. major</i> entre parasitos humanos-----	115
5.1.8	Padrão de expressão das principais enzimas específicas de <i>L. major</i> -----	116
5.2	Busca de Enzimas Específicas em <i>F. oxysporum</i> -----	116
5.2.1	Enzimas Principais-----	118
5.2.1.1	Quitina sintase-----	119
5.2.1.2	Enzima <i>decapping scavenger</i> -----	120
5.2.1.3	Carnitina O-acetiltransferase-----	121
5.2.1.4	3-hidroxantranilato 3,4-dioxigenase-----	121
5.2.1.5	Ureidoglicolato liase-----	122
5.2.1.6	Holocitocromo-c sintase-----	122
5.2.1.7	Esfingomielinase neutra-----	123
5.2.2	Enzimas Auxiliares-----	123
5.2.2.1	2-metilcitrate desidratase-----	123
5.2.2.2	Indoleamina 2,3-dioxigenase-----	124
5.2.2.3	Fosfatidiletanolamina N-metiltransferase-----	124

5.2.2.4	L-metionina (R) -S-oxidase reductase-----	125
5.2.2.5	ATP adenililtransferase-----	125
5.3	Validação das Enzimas Específicas por Análise de balanço de fluxo – FBA-----	126
5.3.1	Análise de balanço de fluxo da reconstrução metabólica de <i>F. oxysporum</i> -----	126
5.3.2	Genes e atividades de enzimas inseridas na Reconstrução metabólica de <i>F. oxysporum</i> -----	130
5.3.2.1	25S rRNA M(1)2142 – metiltransferase-----	131
5.3.2.2	Carboxi-cis,cis-muconato ciclase-----	131
5.3.2.3	Aconitato descarboxilase-----	132
5.3.3	Modelagem tridimensional das duas enzimas análogas consideradas críticas-----	133
6.	PERSPECTIVAS	136
7.	CONCLUSÃO	138
8.	REFERÊNCIAS BIBLIOGRÁFICAS	141
9.	ANEXO 1	201
10.	ANEXO 2	223
11.	ANEXO 3	232

ÍNDICE DE FIGURAS

Figura 1: Fluxograma do processo de depuração de enzimas estritamente específicas.42

Figura 2: Algoritmo de busca homóloga para a classificação de falsos positivos de genes específicos de parasitas em seres humanos. A: As onze possibilidades das associações de proteína Ensembl (S2-E2) que se pode obter com uma região genômica humana que é homóloga (tBLASTn) a uma consulta de proteína parasita (S1-E1). S e E são para o início e o fim de uma homologia de tBLASTn ou coordenadas de genes humanos. A descrição booleana de cada associação entre um hit tBLASTn em uma região do genoma humano e as proteínas Ensembl anotadas nessa região é dada nos lados esquerdo e direito do painel A. '&' é usado aqui em seu sentido booleano, ou seja, um operador lógico AND. Os genes humanos para a proteínas Ensembl que são eventualmente compatíveis com a contraparte homóloga de um parasito (tBLASTn) são modelados por linhas finas. Os genes humanos para proteínas Ensembl na mesma região genômica que a contraparte homóloga de um parasito, mas que não são compatíveis com ele, são representados por linhas tracejadas. B: A árvore de decisão para as associações VERDADEIRO e FALSO dos genes humanos para as proteínas Ensembl com uma região genômica humana que teve um hit homólogo (tBLASTn) com uma proteína de um parasito.-----44

Figura 3: Esquema de pesquisa para falsos positivos de genes específicos de parasitos em seres humanos. O hit homólogo de L. major e proteínas humanas (A) é comparado à estrutura de genes humanos obtida pela pesquisa de tBLASTn (B) e à anotação de CCDS de Havana (C).-----46

Figura 4: Fluxograma do processo de depuração das enzimas estritamente específicas. Nrd significa não-redundante.-----	69
Figura 5: Árvore filogenética de REL1 em Metakinetoplastina obtida por máxima de verossemelhança. Escala: 0,05 substituições por sítio.-----	87
Figura 6: Produção dos componentes de biomassa de <i>F. oxysporum</i> .-----	96
Figura 7: Porcentagem dos componentes de biomassa.-----	96
Figura 8: Modelagem 3D da enzima Pseudouridina sintase (EC:3.5.4.26). A) esquerda: Estrutura monomérica da enzima Pseudouridina sintase (EC: 3.5.4.26) de <i>F. oxysporum</i> (ciano), sobreposta à estrutura de <i>E. coli</i> (APB: 1V9F - laranja). Direito: Gráfico de Ramachandran com porcentagem favorável de 90%. B) Esquerda: potencial eletrostático (Kb.T.ec ⁻¹) - azul: negativo, vermelho: positivo, toda a estrutura. Direito: potencial eletrostático do sítio ativo. C) Hidrofobicidade - laranja: hidrofóbico; azul: hidrofílico. Esquerda: toda a estrutura. Direita: foco do sítio ativo.-----	101
Figura 9: Modelagem 3D da enzima responsável pela biossíntese de Riboflavina (EC: 3.5.4.26). A) Esquerda: Estrutura monomérica da biossíntese de Riboflavina (EC: 3.5.4.26) de <i>A. thaliana</i> (ciano), sobreposta à estrutura de <i>A. baumannii</i> (PDB: 3ZPG - laranja). Direito: Gráfico de Ramachandran com porcentagem favorável de 91%. B) Esquerda: Potencial eletrostático (Kb.T.ec ⁻¹) - azul: negativo, vermelho: positivo, toda a estrutura. Direito: potencial eletrostático do sítio ativo. C) Hidrofobicidade - laranja: hidrofóbico; azul: hidrofílico. Esquerda: toda a estrutura. Direita: foco do sítio ativo.	102

LISTA DE TABELAS

Tabela 1: Fonte de dados dos parasitas humanos importantes para comparação com enzimas estritamente específicas de <i>L. major</i> .-----	54
Tabela 2: Fontes dos dados de genomas completos e proteínas das plantas utilizadas para comparação de similaridade com <i>F. graminearum</i> .-----	60
Tabela 3: Endereço eletrônico utilizados para obter o genoma, as sequências proteicas e genes de <i>A. thaliana</i> , <i>G. max</i> , <i>B. rapa</i> , <i>R. communis</i> e <i>J. curcas</i> .-----	62
Tabela 4: Proteínas de <i>H. sapiens</i> recuperadas do <i>nr</i> utilizando o trecho de DNA do <i>subject</i> correspondente à região homóloga entre as proteínas de <i>L. major</i> e o genoma humano como <i>query</i> .-----	71
Tabela 5: Anotações de números de EC de proteínas humanas homólogas com enzimas de <i>L. major</i> classificadas como específicas funcionais.-----	73
Tabela 6: Enzimas putativas funcionais e estritamente específicas de <i>L. major</i> que dividem ECs.-----	76
Tabela 7: Enzimas específicas de <i>L. major</i> após filtragem com HMM.-----	77
Tabela 8: Proteínas homólogas de <i>T. cruzi</i> , <i>T. brucei</i> e <i>P. falciparum</i> homologas às enzimas específicas estritas de <i>L. major</i> .-----	80
Tabela 9: Relação de paralogia ou analogia entre os grupos de proteínas de <i>L. major</i> .-----	82
Tabela 10: Enzimas específicas cujos genes são únicos no genoma de <i>L. major</i> .-----	84

Tabela 11: Enzimas consideradas específicas a *F. graminearum* (FGR) e *F. oxysporum* cepa Fo5176 (FOX) em relação a *A. thaliana*, *G. max*, *B. rapa*, *R. communis*, *J. curcas*.

-----91

Tabela 12: Novos genes e atividades enzimas inseridas na Reconstrução metabólica de *F. oxysporum*.-----94

Tabela 13: Escores de análise do modelo de *F. oxysporum*.-----100

Tabela 14: Escores de análise do modelo de *A. thaliana*.-----100

PREÂMBULO

Este trabalho concentrou-se principalmente na avaliação de uma metodologia para buscar e avaliar a relevância do desenvolvimento de inibidores para alvos específicos de um organismo patógeno que pudesse ser aplicada em sistemas hospedeiro-parasito em geral.

Em um primeiro passo, avaliamos as fontes de erro em uma ferramenta, AnEnPi, desenvolvida na Fiocruz, com um modelo bem descrito e confiável no plano da anotação genômica; tratou-se do sistema *Homo sapiens* vs. *Leishmania major*. Este modelo permitiu dissecar as fontes de erro no ato de identificar enzimas específicas na *L. major* em relação ao seu hospedeiro humano. Infelizmente, no tempo desta experimentação, não havia uma reconstrução metabólica disponível neste sistema, assim, não foi possível avaliar *in silico* como um método alternativo para entender se os alvos específicos apontados pela investigação em *L. major* eram realmente promissores para afetar o desempenho do parasito ao nível celular.

Para testar a relevância de propor enzimas específicas como alvos para desenvolvimento de inibidores de patógenos em sistemas hospedeiros-parasitos, aplicamos a mesma metodologia de diagnóstico de enzimas específicas desenvolvidas em *H. sapiens* vs. *L. major* em um sistema planta vs. fungo. Plantas e fungos têm aproximadamente o mesmo intervalo de tempo evolutivo desde a divergência deles (1 bilhão de anos) que os vertebrados e protozoários o que faz a comparação dos dois sistemas relevante.

Escolhemos o sistema *Arabidopsis thaliana* vs. *Fusarium oxysporum* pelo fato de *A. thaliana* ser um modelo para biologia molecular em plantas e *F. oxysporum* ser um caso de importância econômica na agricultura.

Pouco é ainda conhecido sobre as micoses humanas e a biologia molecular dos sistemas hospedeiro-parasitas envolvendo humanos e fungos pode aprender muito dos sistemas envolvendo plantas e fungos já que estes sistemas são investigados a décadas em agricultura. No sistema planta vs. fungo, existia uma reconstrução metabólica no tempo desta experimentação, tanto para *A. thaliana* que para *F. oxysporum*, assim, pudemos neste sistema avaliar a relevância das enzimas específicas como alvos para desenvolvimento de inibidores e também testar uma estratégia complementar.

1. INTRODUÇÃO

1.1 Alvos em Sistemas Hospedeiro-Parasitos

O método para fazer com que um fármaco chegue ao mercado é um processo de longo prazo que custa bilhões de dólares. Em 2014, o *Tufts Center for the Study of Drug Development* estimou que o custo associado ao desenvolvimento e ao fornecimento de um novo fármaco ao mercado aumentou quase 150% na última década. O custo agora é estimado em \$ 2.6 bilhões de dólares (1,2).

Durante os processos de descoberta e desenvolvimento de medicamentos, a probabilidade de uma falha na metodologia aplicada é alta assim que 90% dos medicamentos que entram em ensaios clínicos não são aprovados pela *Food and Drug Administration* (FDA) e não chegam ao mercado consumidor. Aproximadamente 75% dos custos são devidos a falhas que ocorrem ao longo da descoberta de fármacos e metodologias de desenho (1,3).

O projeto do genoma humano disponibilizou uma quantidade substancial de dados de sequências que podem ser usados na busca de fármacos. Dentre os desafios da genômica funcional, um dos principais é determinar os papéis dos genes (4). O significado funcional de um gene pode ser estudado aos níveis dos genoma, transcriptoma, proteoma e metaboloma (5,6). No entanto, como muitos genes não parecem ser necessários para a sobrevivência, a produção de energia, os fluxos metabólicos ou o crescimento celular (4,6), a tarefa de determinar a função do gene é muito mais complexa do que a simples determinação da expressão gênica.

Uma descoberta relevante foi que a porcentagem de DNA (menos de 3% do genoma) que codifica proteínas - aproximadamente 20.000 genes humanos - é baixa

(~1,5%). No entanto, análises mais recentes indicam que uma proporção de até 80% do genoma humano é funcional, transcrita, e ligada a proteínas reguladoras, ou associada a outras funções bioquímicas (7).

O alvo de um fármaco é uma biomolécula (s), normalmente é uma proteína ou uma enzima que se apresenta isolada ou complexada (8). Muitas proteínas são consideradas possíveis alvos para intervenções de fármacos que controlam doenças humanas. O número mais recente de alvos de fármacos foi estimado em centenas, com base em uma análise feita antes de 2007 (9,10). No entanto, o número de proteínas que podem funcionar como alvos terapêuticos é substancialmente maior de acordo com o site da base de dados DrugBank (<http://www.drugbank.ca/>). A versão atual deste banco de dados (5.0.9) contém 10.500 entradas de medicamentos que estão ligadas a 4.772 sequências de proteínas não-redundantes (*nr*) (alvo de fármacos/enzima/transportador) ligadas às entradas dos medicamentos (11). Santos et al. (2017) mapearam alvos moleculares entre os quais os fármacos aprovados atuam; neste estudo foi realizada a curadoria de um total de 893 biomoléculas humanas, sendo 667 proteínas do genoma humano provenientes de alvo de fármacos para doenças humanas derivadas de patógenos entre as quais os 1.578 fármacos foram obtidos da lista dos medicamentos aprovados pela FDA – USA (12).

Uma proteína *druggable* é aquela que possui sítio (s) que favorece (m) a interação com pequenas moléculas semelhantes a fármacos, sejam elas de origem endógena ou exógena e, portanto, contém um local de ligação (8,13,14). Espera-se que esses locais de ligação tenham certas características que permitam uma ligação específica ao sítio de alta afinidade pela molécula semelhante ao fármaco. Tal como acontece com todos os alvos de fármacos, um potencial alvo de fármaco proteico deve estar relacionado a um processo de doença (9).

No tratamento de doenças infecciosas, o principal desafio é projetar fármacos seletivos que visam o patógeno sem prejudicar o paciente. Os esforços concentrados para a descoberta de fármacos, através de parcerias público-privadas, para as principais doenças parasitárias de protozoários do mundo em desenvolvimento - malária, leishmaniose e tripanossomíase - renovaram o interesse em desenvolver novos medicamentos e vacinas que possam ser acessíveis para a população afetada, principalmente a população pobre (15).

O processo de descoberta de fármacos se inicia com a busca de alvos que devem preencher, no caso de doenças infecciosas, dois requisitos: (i) serem específicos, ou seja, os alvos não devem apresentar uma contraparte no hospedeiro humano que possa dar origem a efeitos tóxicos (16,17) e (ii) serem essenciais para a sobrevivência do parasito. Um fármaco que seja capaz de inativar uma proteína/enzima específica de um patógeno, provavelmente, é seguro para o seu hospedeiro. Entretanto, um fármaco que seja capaz de inativar uma proteína/enzima que esteja presente tanto no patógeno quanto no hospedeiro poderá causar efeitos tóxicos no hospedeiro.

As proteínas homólogas são aquelas que estão presentes no patógeno e no hospedeiro e compartilham uma história evolutiva comum, ou seja, descendem com divergência de um gene ancestral comum. Na maioria das vezes, as proteínas homólogas possuem estruturas tridimensionais semelhantes. Em contrapartida, as enzimas análogas são o resultado de acontecimentos evolutivos independentes. Por definição, essas enzimas são resultantes da evolução convergente de proteínas independentes, ou seja, elas não se originaram de um ancestral comum (ortólogos) (18). Apesar de possuírem a mesma função enzimática, as enzimas análogas não tem similaridade de sequência primária e normalmente não possuem similaridade estrutural (18,19).

O metabolismo dos patógenos explora fornecimento de energia para sobrevivência e proliferação, o que torna as enzimas metabólicas uma excelente oportunidade para alvos no desenvolvimento de fármacos (17). Contudo, os patógenos eucariotos tendem a ter semelhanças bioquímicas com as células do hospedeiro. Nesses casos, talvez seja necessário deixar de lado o critério da especificidade e considerar os alvos de fármacos que fazem parte de vias que também ocorrem no hospedeiro e que até mesmo possam ser essenciais. Nestes casos, se pode privilegiar uma forma de especificidade baseada em diferença de expressão entre o parasito e o seu hospedeiro para o alvo considerado, na tentativa de minimizar os efeitos colaterais adversos do inibidor para o hospedeiro ou até buscar alvos em vias de sinalização (17).

Uma forma de avaliar a essencialidade de uma enzima alvo é através de experimentos que inibem genes ou experimentos *in silico* usando modelos metabólicos em escala genômica (20) o que permite a identificação de alvos que são essenciais para o patógeno, mas não para o hospedeiro (21).

Em vias de sinalização, vários métodos para avaliação da essencialidade de uma proteína alvo já foram descritos: (i) assinatura de pares de sequência (22), (ii) perfis de domínios proteicos (23,24), e (iii) homologia de sequências (25). Os mais robustos integram uma série de tipos de dados genômicos, como a expressão gênica e o fenótipo *knockout*, e utilizam estruturas sofisticadas de aprendizado de máquina, como redes bayesianas (26), árvores de decisão (27), florestas randômicas e máquinas de vetor de suporte (28) para predizer a interações proteína-proteína.

Além disso, outros tipos de ferramentas para a descoberta de inibidores assistida por computador, ou seja, *in silico*, podem atuar como um atalho virtual auxiliando na expedição deste longo processo e potencialmente reduzindo o custo de pesquisa e

desenvolvimento. As tecnologias de descoberta de inibidores auxiliadas por computador são ferramentas poderosas que podem reduzir o número de ligantes que precisam ser rastreados em ensaios experimentais. Experimentos computacionais não só reduzem o custo associado à descoberta de medicamentos ou de defensivos agrícolas, assegurando que o melhor composto seja utilizado em ensaios *in vitro* e *in vivo*, mas também podem reduzir o tempo necessário para que uma molécula ativa (*lead*) atinja o mercado consumidor (29).

As metodologias baseadas no atracamento virtual de um ligante no alvo proteico assistido por computador podem ser classificadas em dois grupos: baseada em estrutura e baseada em ligante. Os métodos baseados em estrutura requerem que a informação tridimensional (3D) do alvo seja conhecida. Já os métodos baseados em ligantes são usados quando a estrutura 3D do alvo não é conhecida. Eles usam informações sobre as moléculas que se ligam ao alvo de interesse. Os acessos são identificados, filtrados e otimizados para obter potenciais candidatos a medicamentos que serão testados experimentalmente *in vitro* (29).

A descoberta de *leads* baseadas no encaixe *in silico* de inibidores em modelos de estruturas 3D dos alvos proteicos provou ser de grande valor nos processos de desenho racional de moléculas para o controle fisiológico de seres vivos, bem como na triagem de inibidores. O atracamento molecular *in silico* pode permitir uma economia substancial no custo e tempo envolvidos no processo de liberação de medicamentos (30).

1.2 Bioinformática

De acordo com a definição do dicionário *Oxford English Dictionary*, a bioinformática conceitualiza a biologia em termos de moléculas (no sentido da química física) e aplica "técnicas de informática" (derivadas de disciplinas tais como matemática aplicada, ciência da computação e estatística) para entender e organizar a informação associada a essas moléculas, em grande escala. Em resumo, a bioinformática é um sistema de informação gerencial para biologia molecular e tem muitas aplicações práticas.

No início dos anos 1980, os métodos de sequenciamento de DNA se tornaram amplamente disponíveis para a comunidade científica, logo houve um crescimento exponencial de dados moleculares de sequências. Como consequência, os resultados gerados se acumularam em diversos bancos de dados e métodos computacionais, como algoritmos que visavam a busca por similaridade de sequências assim que predição de estruturas e função destas, foram desenvolvidos para que estes dados pudessem ser recuperados e analisados (31).

A grande quantidade de dados que foi gerada levou à adoção precoce de ferramentas para análises automatizadas não supervisionadas de dados biológicos em meados da década de 1990 dando origem a um novo campo, a bioinformática, que combina elementos de biologia e ciência da computação (32–34).

Dentre os objetivos da bioinformática estão os (i) de organizar os dados de forma a permitir que os pesquisadores acessem informações existentes e apresentem novas entradas à medida que são produzidas; (ii) de desenvolver ferramentas e recursos para

ajudar na análise de dados; (iii) de usar ferramentas disponíveis para analisar os dados e e (iv) interpretar os resultados de forma biologicamente significativa (34).

1.2.1 Genômica Funcional

De acordo com a definição proposta por Benner e Gaucher (2001) a genômica funcional é o campo que utiliza ferramentas computacionais para inferir a função biológica a partir de sequências genômicas suplementadas com experimentos conduzidos, em parte, por hipóteses geradas a partir da análise de sequência genômica (35). Em outras palavras, ela reúne dados de transcritos de genes, proteínas e metabólitos para responder a questão colocada por todos os projetos de análise de sequenciamento do genoma: qual é a função biológica de cada gene? (36).

Dentre as ferramentas utilizadas em genômica funcional, a ferramenta AnEnPi (37) (*Analogous Enzymes Pipeline*, <http://anenpi.fiocruz.br/>), utilizando sequências armazenadas no banco de dados do *Kyoto Encyclopedia of Genes and Genomes* (38) (KEGG), implementa uma série de algoritmos que executam as seguintes tarefas: (i) o agrupamento de sequências de proteínas por homologia utilizando o *Basic Local Alignment Search Tool* (39) (BLAST); (ii) a classificação de sequências em diferentes grupos homólogos como enzimas análogas quando têm a mesma função enzimática, isto é, *Enzyme Commission Number* (18) (EC); (iii) a detecção de enzimas específicas; (iv) a anotação de funções de proteínas usando BLAST ou HMMER (40); e (v) a geração de mapas metabólicos usando as ferramentas fornecidas pelo KEGG PATHWAY.

1.3 Seleção do Modelo Hospedeiro – Parasito

No Brasil, nove das dez doenças tropicais negligenciadas estabelecidas pela Organização Mundial da Saúde (OMS) estão presentes no país (41). Diante das inúmeras doenças infecciosas que assolam a população mundial, a Leishmaniose é considerada um grave problema de saúde. Esta doença é classificada com base no painel do TDR *Research and Training in Tropical Diseases* (TDR) / Organização Mundial da Saúde na categoria um que inclui as principais doenças tropicais negligenciadas que não estão sob controle e estão emergindo (42).

A relevância da leishmaniose deve-se à (i) rápida urbanização e expansão de sua área de transmissão, (ii) a elevada taxa de letalidade com cerca de 10,0% dos pacientes infectados passando ao óbito (43), (iii) e das dificuldades encontradas para seu controle que estão relacionadas à necessidade do diagnóstico e tratamento precoce dos casos humanos, principalmente no caso da leishmaniose visceral (LV), para evitar óbitos (43,44).

Uma vez que o número de tecnologias ômicas (*omics*) estão aumentando, estas estão sendo aplicadas às investigações de determinantes do fenótipo da leishmaniose (45), do modo de ação dos medicamentos atuais (46), e da biologia do parasito envolvido (47). Esses estudos melhoraram a compreensão da patogênese da doença em humanos e dos possíveis mecanismos de resistência a fármacos anti-leishmaniose. Sem dúvida, as abordagens ômicas estão revelando detalhes das relações íntimas entre hospedeiros, parasitas e vetores. O desenvolvimento de novas ferramentas que sejam capazes de refinar os conhecimentos promoverá o desenvolvimento de novos meios de controle (por exemplo, vacinas antivectores) que poderiam auxiliar a luta contra leishmaniose (48).

1.3.1 Modelo *H. sapiens* – *L. major*

1.3.1.1 *H. sapiens*

O Projeto Genoma Humano (49–51) foi provavelmente um dos empreendimentos científicos mais ambiciosos até o momento e providenciou as primeiras peças para desenvolver modelos para prever respostas de células a desafios específicos e, em particular, doenças (52). Porém, apesar do imenso progresso desde o sequenciamento do primeiro genoma humano há mais de 10 anos, há ainda muito que aprender sobre o genoma humano (53,54).

Apenas 1,5% do genoma humano, cerca de 45 Mbp, codifica as sequências de proteínas de genes clássicos, sendo que existem aproximadamente 22.000 genes no genoma humano, a maioria codificando uma proteína. No entanto, é possível observar que aparentemente existem milhares de genes que não codificam proteínas, mas possuem uma função no RNA transcrito e processado, por exemplo, genes microRNA (miRNA), genes RNA ribossômicos e RNA não-codificadores longos intergênicos (55).

O transcriptoma é definido como o conjunto de todas as moléculas de RNA que estão presentes em uma determinada célula em um determinado momento. Estima-se que existem aproximadamente 300.000 moléculas de mRNA numa célula. Dos 22.000 genes existentes em nosso genoma, somente de 6.000 a 8.000 são expressos em níveis significantes em células diferenciadas (56).

O exoma é definido como a sequência de DNA combinada de todos os éxons de genes codificadores de proteínas e RNA no genoma. Diferentemente do transcriptoma que compreende todas as moléculas de RNA em uma determinada célula, de modo que

varia entre tipos celulares, o exoma é idêntico para todas as células de um organismo (57).

Com o desenvolvimento e inserção de técnicas de sequenciamento de alta vazão para DNA e RNA, a análise dos genomas e transcriptomas tem sido possível a uma velocidade e custo razoáveis. Isto é em parte devido à natureza do DNA e RNA, que são constituídos por moléculas de apenas quatro componentes, os nucleótidos, que são facilmente amplificáveis pelas polimerases. No entanto, em relação à análise de proteomas, estas são mais delicadas, pois os 20 aminoácidos possuem propriedades físico-químicas bastante diferentes. Além disso, a abundância de proteína pode diferir em ordens de dez de grandeza em células e as modificações pós-translacionais de proteínas, por exemplo, fosforilação e glicosilação, e interações proteína-proteína são altamente dinâmicas no espaço e no tempo (52).

1.3.1.2 *L. major*

A leishmaniose é uma doença transmitida pela picada de fêmeas de flebotomíneos (Diptera: Psychodidae: Phlebotominae) e causada por parasitos protozoários obrigatórios pertencentes ao gênero *Leishmania* (*Trypanosomatidae*). Atualmente, existem 350 milhões de pessoas em áreas de risco e 12 milhões de casos de infecções. As áreas de risco compreendem os locais onde a doença é endêmica como áreas dos trópicos, subtropicais e da bacia do Mediterrâneo, incluindo mais de 98 países (58,59). Entre as regiões endêmicas nos cinco continentes, há uma incidência estimada de 0,7-1,2 milhões de casos de leishmaniose cutânea (LC) e 0,2-0,4 milhões de casos de LV nesses países (58). Entretanto, a doença se encontra ausente na Nova Zelândia e no Pacífico Sul.

São considerados hospedeiros vertebrados, animais como canídeos, roedores, marsupiais, dassies e seres humanos. Existem aproximadamente 53 espécies de *Leishmania*, isso sem considerar os sinônimos, nos cinco subgêneros e complexos: *Leishmania*, *Viannia*, *SauroLeishmania*, complexo *L. enriettii*, e *ParaLeishmania*; sendo que destas 31 espécies são conhecidas por serem parasitos de mamíferos e 20 espécies são patogênicas de humanos (59).

Em humanos, de acordo com a espécie dos parasitos, *Leishmania* pode causar três tipos diferentes de manifestação da doença: LV, LC, e leishmaniose mucocutânea (LMC). Existem outras formas da doença que incluem leishmaniose cutânea difusa (LCD) e leishmaniose dérmica pós-calazar (LDPC) que muitas vezes estão ligadas ao status de imunidade do hospedeiro (60). A forma mais comum da doença é a leishmaniose cutânea e Afeganistão, Argélia, Colômbia, Brasil, Irã, Síria, Etiópia, Sudão do Norte, Costa Rica e Peru representam juntos 70% a 75% da incidência global estimada de casos de LC. Em relação aos casos de leishmaniose visceral, Índia, Bangladesh, Sudão, Sudão do Sul, Brasil e Etiópia detêm mais de 90% do total de todos os casos (58).

Os antimoniais pentavalentes são a primeira linha de tratamento padrão, mas o surgimento de resistência e toxicidade a eles limitou sua utilidade (61). Os medicamentos incluem: (i) anfotericina B na sua forma livre ou lipossomal, uma combinação de antimoniais pentavalentes e paromomicina, e miltefosina, para a forma visceral e (ii) paromomicina, fluconazol ou pentamidina, para a forma cutânea (62).

Em relação ao mecanismo de ação, a anfotericina B na formulação lipossomal se liga ao ergosterol, que é o principal esterol em membranas celulares fúngicas e de *Leishmania* (63), o que resulta em uma mudança na permeabilidade da membrana,

promovendo vazamentos de íons monovalentes, distúrbios metabólicos e morte celular. O medicamento AmBisome, formulação comercial, é caro, o que dificulta o uso em grande escala nos países em desenvolvimento.

O efeito dos antimoniais pentavalentes é pleiotrópico, pois eles podem atuar (i) como um pró-fármaco que se converte em antimônio trivalente ativo e mais tóxico, (ii) diretamente em alvos moleculares no metabolismo redox do tiol (64), como a tripanotona e a glutatona (65), e (iii) através de tióis e ribonucleósidos (66).

A paromomicina é um antibiótico que foi proposto para sucessivamente (i) se unir às proteínas da haste paraflagelar e à proibitina, (ii) ser internalizada por endocitose, e (iii) interagir com uma adenosina trifosfatase (ATPase) tipo P (H^+) (67).

A miltefosina é um fosfolípido de alquilo cujo mecanismo de ação é inibir o metabolismo dos fosfolípidios, diminuindo a fosfatidilcolina e aumentando a fosfo-etanol-etanolamina (68) com a consequência da indução da morte celular por apoptose (69).

A anfotericina B, os antimoniais pentavalentes, paromomicina e miltefosina induzem efeitos colaterais nocivos e possuem eficácia variável no tratamento da leishmaniose, dependendo da localização geográfica. Um tratamento combinado tem as potenciais vantagens de reduzir a duração do tratamento, reduzir a dose global de medicamentos e reduzir a probabilidade de seleção de parasitas resistentes aos medicamentos (70).

As infecções causadas por *Leishmania* estão se tornando grandes problemas de saúde pública em escala global. Muitas espécies de *Leishmania* em todo o mundo estão obtendo elevados níveis de resistência, conforme estimado pelo *World Health*

Organization (WHO) (70). O arsenal de fármacos disponíveis para o tratamento de infecções por *Leishmania* é limitado e existem alguns outros fármacos em vários estágios do processo de desenvolvimento (71). Entretanto, as espécies de *Leishmania* que possuem resistência são relativamente difíceis de observar e manter em ambientes de laboratório (72).

Como o repertório atual de fármacos é limitado e o aumento da resistência contra os medicamentos tem sido uma grande preocupação, é necessário identificar novos alvos para o desenvolvimento de novos fármacos. As sequências do genoma de *L. major* revelaram imensa quantidade de informação e deram a oportunidade de buscar alvos de inibidores exclusivos desse parasito. A utilização da informação para criar uma molécula de fármaco candidato requer a combinação de toda a tecnologia e a utilização de uma abordagem multidisciplinar, desde a caracterização da proteína alvo até a triagem de compostos (61).

1.3.2 *H. sapiens* – *Leishmania*: interação hospedeiro – parasito

Os parasitos pertencentes ao gênero *Leishmania* executam os seus ciclos de vida em dois hospedeiros: flebotomíneos e mamíferos. Quando um flebotomíneo infectado faz seu repasto sanguíneo em mamíferos, ele injeta as formas promastigotas na pele do hospedeiro. E essas formas promastigotas são fagocitadas por células fagocíticas, como macrófagos, neutrófilos e células dendríticas (CD). Dentro dos macrófagos, as formas promastigotas se diferenciam em amastigotas, que são as formas proliferativas de *Leishmania*. Os parasitos na forma amastigota proliferam nos macrófagos do tecido e migram para outros macrófagos dependendo de fatores bioquímicos do parasito e do hospedeiro (73). Na LC, a infecção é normalmente limitada à pele e ao sistema linfático,

porém esta pode evoluir para a infecção de tecidos mais profundo, caracterizando a LCD ou ainda invadir as mucosas levando ao quadro de LMC. O ciclo de vida só é considerado completo quando o mosquito se alimenta perto das lesões da pele e durante seu repasto suga as formas amastigotas que vão para o seu intestino médio, onde posteriormente se diferenciam em formas promastigotas (74–77).

Assim como em outras infecções, a resposta imune à LC depende de muitos fatores do hospedeiro, bem como de diferenças entre as espécies infectantes de *Leishmania* (77). Em camundongos, infecções experimentais exibem espectro de apresentações clínicas dependendo da cepa do camundongo e da espécie infecciosa do parasito ou da cepa do parasito que foi utilizada (78–85).

Em pacientes com leishmaniose, o espectro imunológico observado varia desde indivíduos com uma resposta forte de células T, caracterizada por uma hipersensibilidade do tipo tardia (HTT) e altos níveis de interferon- γ (IFN- γ), a indivíduos que não possuem resposta HTT, mas podem apresentar níveis elevados de anticorpos (86).

Leishmania spp. são mortos por macrófagos ativados pelo IFN- γ e não são neutralizados por anticorpos, os indivíduos que apresentam uma HTT forte têm poucos parasitos em suas lesões, enquanto aqueles com apenas uma resposta humoral são incapazes de controlar a carga parasitária (86,87). Logo, pacientes ausentes de uma resposta de células T apresentam a forma LCD como consequência da doença. Além disso, pacientes com uma resposta imune exagerada também desenvolvem um outro fenótipo de doença grave conhecido como LMC, que é impulsionada pela imunopatologia. E entre estes extremos estão os pacientes que desenvolvem lesões que

podem se auto curar ou tornar-se crônicas, com níveis intermediários de respostas de células T e anticorpos (88).

Para o controle de qualquer doença infecciosa, um meio que pode ser utilizado é a vacinação. Como outras doenças infecciosas, a leishmaniose também deve ser controlada pela vacinação. Entretanto, atualmente não há vacinas aprovadas para uso geral; no entanto, várias pesquisas e desenvolvimentos neste sentido ainda se encontram em andamento. Até o momento, foram propostas numerosas tentativas de desenvolvimento de uma vacina bem-sucedida contra a leishmaniose e existem vários tipos de vacinas candidatas, mas principalmente profiláticas. As vacinas para a profilaxia da leishmaniose, que estão em fase de desenvolvimento, podem ser divididas em três categorias: (i) vacinas vivas atenuadas de *Leishmania*, incluindo novas cepas geneticamente modificadas; (ii) vacinas contendo os parasitos mortos, constituídas pela *Leishmania* inteira morte ou por frações do parasito; e (iii) vacinas definidas, por exemplo, proteínas recombinantes, vacinas de DNA (89).

Embora nos últimos anos um grande grupo de pesquisadores tenha concentrado seus esforços nesta questão, apenas três vacinas candidatas foram para ensaios clínicos, até o momento: (i) vacina contendo os parasitos mortos no Brasil para imunoterapia humana; (ii) vacina com parasitos vivos, porém atenuados para humanos no Uzbequistão; e (iii) vacina de segunda geração para profilaxia de cães no Brasil (90).

Como as vacinas eficazes contra a leishmaniose estão ainda em desenvolvimento, as medidas de controle atuais dependem exclusivamente da quimioterapia para tratar a população. E uma vez que os antimoniais são nocivos e apresentam elevada resistência, é necessário identificar novos alvos de fármacos que sejam exclusivos desses parasitos. A nossa estratégia foi de buscar alvos principalmente em caminhos metabólicos que

sejam essenciais e que possam ser usados para o desenvolvimento de fármacos em *Leishmania*.

1.4 Seleção Modelo Planta – Fungo

A planta *Arabidopsis thaliana* é uma dicotiledônea que é utilizada como modelo para diversas pesquisas científicas (91). Devido a este fato, existem diversos bancos de dados online contendo informações relacionadas a dados genéticos e de biologia molecular da planta, entre eles o *The Arabidopsis Information Resource (TAIR)* (92). Devido ao seu genoma de tamanho pequeno, facilidade experimental e tempo de geração curto *A. thaliana* foi escolhida como organismo modelo para pesquisas em biologia molecular e genética de plantas (93,94).

Os fungos causam doenças em plantas que destroem e contaminam a cada ano uma significativa proporção da produção agrícola global, o que os tornam a classe mais agressiva de patógenos de plantas (95,96). O *F. oxysporum* é um fungo transmitido pelo solo agente causal da murcha vascular, uma doença que afeta uma grande variedade de culturas economicamente importantes em todo o mundo (97). Além de sua atividade bem estudada como patógeno vegetal, *F. oxysporum* é conhecido como um importante patógeno emergente de humanos devido ao número crescente de casos graves relatados e à sua ampla resistência aos medicamentos antifúngicos disponíveis (98,99).

De acordo com Vartivarian et al. (1993) (100); Ponton et al. (2000) (101); Nucci e Anaissie (2002) (102) a fusariose humana foi a causa mais frequente de infecções fúngicas invasivas (sistêmicas) em pacientes imunocomprometidos, frequentemente com resultados letais na década de 2000 (103). Entretanto, as infecções causadas podem

variar de uma invasão superficial ou localmente disseminada, dependendo do estado imunológico do hospedeiro (104).

Em um estudo realizado em 9 centros de transplantes, fungemia e manifestações cutâneas foram relatadas em 44% e 75% dos casos de fusariose, respectivamente (105). As espécies mais frequentes que são associadas à causa de doenças em humanos são *F. solani*, *F. oxysporum*, e *F. moniliforme* (105–107). Sendo que as espécies de *Fusarium* parecem ser resistentes *in vitro* durante a terapia com anfotericina B em novas infecções; as espécies de *Fusarium* também parecem ter resistência *in vitro* aos triazoles (105). A taxa de mortalidade dos pacientes geralmente excede 80% aos 90 dias após o início da infecção (105,107).

Em pacientes oncológicos e hematológicos, especialmente aqueles com períodos prolongados de neutropenia, a doença pode ser recorrente (108). Porém, não há consenso quanto ao melhor tratamento (109). Contudo a partir da década de 1990, a ocorrência de fusariose em pacientes infantis e adultos oncológicos tem sido bastante relatada (98,110–117).

Dada a dupla capacidade de causar doenças tanto nas plantas quanto nos seres humanos, *F. oxysporum* poderia servir de modelo universal para o estudo dos mecanismos de virulência fúngica (118,119).

1.4.1 Modelo *A. thaliana* – *F. oxysporum*

1.4.1.1 *A. thaliana*

O gênero *Arabidopsis* compreende nove espécies de plantas de pequeno porte pertencentes à família Brassicaceae nativas da Eurásia, África do Norte e América do Norte (120,121). Como espécies do gênero *Arabidopsis* podemos citar: *Arabidopsis arenosa*, *Arabidopsis cebennensis*, *Arabidopsis croatica*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Arabidopsis neglecta*, *Arabidopsis pedemontana*, *Arabidopsis suecica* e *A. thaliana*.

Dentre as espécies pertencentes ao gênero, *A. thaliana* tornou-se o sistema modelo de planta preeminente em biologia vegetal por causa de seu genoma pequeno, ciclo de vida curto, biologia prática e acessibilidade à análise genética (122–124). Ainda, pelo fato de ter menor genoma conhecido em angiosperma, tornou-se o organismo de eleição para uma ampla gama de estudos em plantas (125), não só para a biologia da planta, mas também para abordar questões fundamentais de estrutura e função biológica comum a todos os eucariontes (126).

É proposto que *A. thaliana* tenha sido originada na Ásia Central e se propagou por toda a Eurásia. Durante a última glaciação, *A. thaliana* foi confinada ao limite sul do seu centro de origem, e depois do recuo do gelo, grande parte da Europa foi recolonizada por diferentes populações, resultando em padrões de mistura complexos. Hoje, *A. thaliana* está presente em todo o hemisfério norte, principalmente em regiões de clima temperado, desde as montanhas do Norte da África até o Círculo Ártico. Como muitas outras plantas Europeias, ela também invadiu a América do Norte, a maioria provavelmente durante tempos históricos (125,127–131).

A ascendência de *A. thaliana* em se tornar uma das espécies mais populares na investigação básica de plantas (132), apesar de sua falta de valor econômico, é devido à sua genética favorável e ao seu ciclo reprodutivo de apenas três semanas. A planta possui um genoma diplóide de apenas 125 a 150 Mb distribuído ao longo de cinco cromossomos, com cerca de 26.000 genes codificadores de proteínas (*The Arabidopsis Genome Initiative*, 2000). A facilidade com que ela pode ser transformada de forma estável excede a de qualquer outra angiosperma (93).

A. thaliana foi a primeira espécie de plantas com genoma sequenciado. Esta sequência inicial era de uma única estirpe homocigota (acesso) e foi de qualidade muito alta com cada cromossomo representado por apenas dois *contigs*, um para cada braço (*The Arabidopsis Genome Initiative*, 2000). Além de análises funcionais, os 120 Mb da sequência de referência do acesso da Columbia (Col-0) provou ser um benefício para genética evolutiva e ecológica. Uma vantagem particular a este respeito é que a espécie é auto-fertilizada, e a maioria das cepas coletadas na natureza são homocigotas para todo o genoma. Isto distingue *A. thaliana* de outros organismos modelo, como o rato ou a mosca da fruta. Nestes sistemas, linhagens puras foram derivadas, mas elas não representam quaisquer indivíduos realmente encontrados na natureza (133).

As plantas com flores apareceram cerca de 100 milhões de anos atrás e a anatomia floral em todas elas é relacionada e derivada de estruturas foliares. Na verdade, a comparação do arroz com *A. thaliana* permitiu mostrar que os principais aspectos da fisiologia da floração são altamente conservados entre monocotiledôneas e dicotiledôneas (134).

As plantas em seu ambiente estão constantemente expostas a uma infinidade de micróbios, incluindo aqueles que possuem um estilo de vida patogênico. Apesar desses

encontros contínuos com possíveis pragas microbianas, a colonização bem-sucedida de plantas por microrganismos patogênicos em *habitats* naturais geralmente é rara. A falha que muitos parasitos em potencial enfrentam na tentativa de colonizar diferentes espécies de plantas demonstra a robustez dos sistemas imunológicos das plantas, cujos mecanismos moleculares subjacentes foram largamente decifrados ao longo dos anos (135,136).

Uma forma de defesa que a planta possui contra patógenos é através da biossíntese de metabólitos secundários de baixo peso molecular que contribuem para a restrição do progresso da infecção (137,138). Os metabólitos secundários produzidos pelas plantas constituem um grupo grande de compostos estruturalmente diversificados que podem ser produzidos *in planta* a partir de vários metabólitos primários ou a partir de seus intermediários biossintéticos, constitutivamente ou em resposta a diferentes estímulos ambientais (139,140).

Os metabólitos secundários têm sido sugeridos de forma a contribuir para as interações de plantas com outros organismos (141), e seus potenciais envolvimento nas respostas das plantas aos agentes patogênicos microbianos constitui as respostas imunes das plantas mais estudadas (142). Entretanto, eles possuem distribuições restritas no reino vegetal e não desempenham papéis considerados essenciais no crescimento e no desenvolvimento da planta (143).

A planta *A. thaliana* produz um grande número de metabólitos secundários, incluindo representantes de muitas das principais classes de metabólitos secundários encontrados no reino vegetal. Um total de 170 metabólitos secundários foi agrupado em sete principais classes: (i) Compostos de indol e indol-enxofre; (ii) Glucosinolatos; (iii)

Fenilpropanóides; (iv) Benzenóides; (v) Flavonoides; (vi) Terpenos; (vii) Derivados de ácidos graxos (143).

Em *A. thaliana* e em outras plantas pertencentes à família Brassicaceae, o grupo de metabólitos secundários que normalmente é considerado o mais característico é o glucosinolato (144). Os terpenos constituem a maior classe de metabólitos secundários produzidos pelas plantas no reino vegetal, sendo que *A. thaliana* produz muitos metabólitos secundários pertencentes a este grupo, pois possui 32 genes correspondentes à enzima terpeno sintase (145,146). Os compostos de benzenóides são derivados de fenilpropanóides, os quais são a principal fonte de metabólitos secundários aromáticos nas plantas (143). A classe dos flavonóides representa os membros mais comuns dos policétides, que consistem de um grupo grande de metabólitos secundários muito típicos de fungos e outros microrganismos. Em *A. thaliana*, os principais compostos da classe dos flavonóides são os flavonóis (147,148), que protegem contra a radiação UV-B (149–151). Ainda, *A. thaliana* produz uma série de compostos indol e indol-enxofre. Cinco dessas estruturas de compostos consistem em derivados de triptofano, sendo estes presentes nas frações solúveis ou de paredes celulares de raízes e folhas ou ainda, em exsudados radiculares (152–155). As plantas produzem uma variedade de derivados de ácidos graxos, entretanto, alguns compostos não possuem função associada ao crescimento e desenvolvimento da planta e com isso podem ser associados como metabólitos secundários. Em *A. thaliana*, podemos citar o cicloheptadieno e aldeídos de cadeia longa como exemplos deste caso (156).

1.4.1.2 *F. oxysporum*

O gênero *Fusarium* contém algumas das espécies de patógenos de plantas mais importantes que afetam economicamente e socialmente a agricultura e a horticultura a nível mundial consistindo em um dos patógenos de plantas fúngicas mais intensivamente estudados em diferentes aspectos: patogenicidade, genética populacional, evolução e genômica (157–159). Este gênero possui característica cosmopolita e é composto de fungos filamentosos pertencentes ao grupo dos ascomicetos (Sordariomycetes: Hypocreales: Nectriaceae), com distribuição mundial que inclui 300 espécies filogeneticamente diagnosticáveis (160), com diversos agentes produtoras de toxinas, patogênicos de plantas de importância agrícola. Uma das espécies mais relevantes do gênero que é responsável por causar murcha vascular e podridão radicular em mais de 100 diferentes espécies de plantas de áreas tropicais e subtropicais é *F. oxysporum* (161). Dentre as plantas que são acometidas por fungos pertencentes ao gênero *Fusarium*, podemos citar: *Arabidopsis thaliana*, *Brassica rapa*, *Cajanus cajan*, *Carica papaya*, *Glycine max*, *Gossypium raimondii*, *Jatropha curcas*, *Linum usitatissimum*, *Lotus japonicus*, *Medicago truncatula*, *Populus trichocarpa*, *Prunus persica*, *Ricinus communis*, *Solanum lycopersicum*, *Theobroma cacao* (162).

Por causa da co-evolução do parasito com seus hospedeiros favorecendo o desenvolvimento de características específicas dos fungos na interação parasito-hospedeiro perceptíveis em nível genômico, as espécies de *F. oxysporum* são subdivididas em *formae speciales* com base na sintomatologia, epidemiologia e susceptibilidade de cultivo (163). Os isolados foram divididos em mais de 120 diferentes *formae speciales* de acordo com a sua gama de hospedeiros (164,165). As *formae speciales* ainda podem ser subdivididas em raças fisiológicas mostrando um padrão característico de virulência em cultivares de hospedeiros diferenciais.

Menos de 20% das espécies de *Fusarium* têm um ciclo sexuado conhecido. *F. oxysporum* é uma espécie anamórfica caracterizada por uma série de critérios morfológicos, incluindo forma de macroconídios, estrutura de microconidióforos e formação e disposição de clamidosporos. *F. oxysporum*, por exemplo, inicialmente penetra nas raízes sem promover sintomas e, posteriormente, coloniza o tecido vascular e provoca murchidão maciça, necrose e clorose na parte aérea da planta (166). Após a germinação de conídios e antes da penetração na raiz, o fungo *F. oxysporum* forma um micélio compacto sobre a superfície das raízes (167), provavelmente através de hidrofobinas de superfície, uma classe de proteínas que tem sido implicada na aderência de superfície fúngica às raízes (168). A aderência do fungo à raiz é um passo crítico na interação hospedeiro-parasito uma vez que fungos não patogênicos saprófitos aderem mal às raízes e não penetram a rizoderme (169). O processo de penetração ocorre preferencialmente na zona meristemática, diminui rapidamente de 40% para menor taxa na zona de alongamento assim que na zona pilosa da raiz, e para zero na zona de ramificação da raiz (167). Os exsudatos de raízes das plantas, na zona meristemática parecem favorecer o crescimento do fungo e induzir a formação de hifas de penetração. A formação de hifas de penetração, por sua vez, induz a aposições de paredes em células. Utilizando técnicas de imunologia e de microscopia eletrônica, carboidratos da parede celular de planta e glicoproteínas foram detectados nessas aposições (167). No tomate, a penetração das raízes por *F. oxysporum* ocorre sem a necessidade de estruturas de infecção diferenciadas (170,171) através de aberturas naturais localizadas nas junções das células da epiderme (172) e podem ser livremente classificadas como hemibiotróficas, pois a infecção inicialmente se assemelha a de um patógeno que se baseia em um hospedeiro vivo (biotrófica), mas eventualmente sofre transições para matar e consumir células hospedeiras (necrotróficas) (173).

Os metabolitos tóxicos de *Fusarium* contribuem para a virulência nas plantas; a fumonisina B1 foi apresentada agindo como um desencadeador potente da morte celular programada em plantas, um mecanismo geralmente associado com a defesa contra agentes patogênicos (174). Os tricotecenos são micotoxinas de *F. oxysporum* que provavelmente agem através da inibição da tradução de proteínas e possuem atividades eliciadoras que estimulam a defesa da planta e promovem a morte celular vegetal (175). Enzimas degradadoras de paredes celulares podem ser importantes em, pelo menos, duas fases fundamentais do desenvolvimento da doença da murchidão vascular; primeiramente, durante a penetração das diferentes camadas do córtex da raiz para obter acesso ao sistema vascular e, em segundo lugar, durante a colonização do hospedeiro por propagação ascendente através dos vasos do xilema. *F. oxysporum* sintetiza uma sequência definida de enzimas de degradação da parede celular extracelular, quando entra em contato com a parede da célula hospedeira (176), que são reguladas através da indução do substrato e repressão catabólica (177). A principal endopoligalacturonase produzida por *F. oxysporum* é a poligalacturonase-1 (PG1) (178). Além disso, duas exopoligalacturonase, PG2 e PG3, também foram purificadas (179,180). Todas estas enzimas são secretadas durante o crescimento do fungo no tecido vascular do tomate. Proteases extracelulares foram identificadas em uma série de fungos patogênicos de plantas e podem desempenhar um papel crítico na colonização bem-sucedida do hospedeiro.

À medida que o fungo tenta se estabelecer dentro do hospedeiro, ele tem que lidar com a presença de metabólitos secundários com atividade antimicrobiana, que são parte das defesas da planta (181). Uma colonização bem-sucedida de plantas hospedeiras por *F. oxysporum* leva à necrose dos tecidos infectados, colapso subsequente de vasos vasculares e morte da planta. A resistência de plantas a *F. oxysporum* parece ser

geneticamente complexa e confere dificuldade pelos cruzamentos, portanto, há muito interesse em determinar as bases moleculares e genéticas da imunidade inata da planta contra este tipo de patógenos (182).

Metabólitos secundários são moléculas orgânicas de baixo peso molecular que não são essenciais para o crescimento normal, mas podem proporcionar uma vantagem seletiva em certos ambientes. Os metabólitos secundários de fungos são diversos na estrutura e atividade biológica. Alguns são corantes, outros são tóxicos para as plantas e/ou animais, alguns regulam o crescimento vegetal ou possuem propriedades farmacêuticas (por exemplo, antibióticos), e outros, incluindo alguns metabólitos de *Fusarium* (por exemplo, eniatinas e tricotecenos), contribuem para patogênese da planta (166).

Os fungos são também capazes de sintetizar uma grande variedade de metabólitos, incluindo aminoácidos, peptídeos pequenos, pigmentos e outros produtos naturais com atividades tóxicas potentes, tais como antibióticos e micotoxinas (183,184). As vias que geram metabólitos primários e secundários não são mutuamente exclusivas (185). Talvez de maneira mais informativa, as vias podem ser divididas entre aquelas que são compartilhadas pela maioria dos organismos, o que pode ser considerado como pertencente ao metabolismo geral, e essas vias especializadas que evoluíram em resposta às restrições ecológicas específicas de determinadas linhagens, como resultado, são distribuídas de forma mais restrita taxonomicamente. Os genes constitutivos das vias metabólicas específicas em fungos são muitas vezes ligados fisicamente nos cromossomos que formam agrupamentos de genes (186,187). Grupos de genes do metabolismo dos fungos são distintos dos grupos de genes de desenvolvimento, tipicamente encontrados em genomas de animais, tais como os agrupamentos de genes Hox; enquanto agrupamentos de genes animais são compostos de genes duplicados em

tandem (188), agrupamentos de genes metabólicos de fungos compreendem genes que são evolutivamente independentes. Grupos de genes metabólicos de fungos participam em diversas atividades, incluindo o metabolismo de nitrogênio (189), de carboidratos (190), e vitaminas (187) bem como no catabolismo de xenobióticos (185) e a biossíntese de metabolitos secundários (191).

A duplicação de genes (DG), uma fonte importante de inovação gênica, é frequentemente implicada na evolução do metabolismo fúngico (192,193), especialmente no contexto de duplicação de todo o genoma (194,195) e expansão de família de genes (196,197). Uma segunda fonte fundamental de inovação metabólica em fungos é através da transferência horizontal de genes (HGT) (198–201) casos significativos incluem a transferência de genes envolvidos no catabolismo de xenobióticos (202,203), na produção de toxina (202,204). Mais recentemente, HGT tem sido mostrado ser responsável pela transferência de grupos de genes metabólicos inteiros entre fungos não relacionados (205–212). DG é o processo dominante subjacente à diversidade metabólica dos fungos, enquanto HGT corresponde apenas a ~5% e seu impacto é mais acentuado em linhagens específicas ou categorias metabólicas (213).

Os agrupamentos de genes de biossíntese de metabolitos secundários podem codificar para (i) transportadores que movem os metabolitos através das membranas e (ii) fatores de transcrição (FT) de vias específicas que ativam a expressão de determinados genes. A maioria dos metabolitos secundários de fungos é sintetizada através das atividades de sintetases de peptídeos não ribossomais (*non ribosomal peptide* - SPNR), sintases de policétido (SPK), ou terpeno-sintases (TerpS). Estas enzimas catalisam a condensação ou rearranjo das moléculas estruturalmente simples para formar estruturas mais complexas: peptídeos não ribossomais, policétidos ou

terpenos. A análise do genoma de *F. graminearum* revelou 16 SPK, 19 SPNR e 8 TerpS, porém foi relatado que o fungo era capaz de produzir oito famílias metabólicas secundárias: aurofusarina, butenolide, clamidosporol, culmorina, ciclonerodiol, fusarinas, tricotecenos e zearalenonas (214). De longe o exemplo mais bem estudado de fitotoxinas de *Fusarium* spp. constitui os tricotecenos, sesquiterpenóides que atuam como potentes inibidores de síntese de proteínas eucarióticas, sintetizados por um conjunto de 23 kb (215).

A análise de sequências do genoma de várias espécies de *Fusarium* spp. assim como PCR e pesquisas de *Southern blot*, indicaram diferenças marcantes na distribuição de genes de biossíntese de metabólitos secundários e, portanto, diferenças no potencial genético das espécies referentes em produzir estes metabólitos (162,216–218). Por exemplo, a distribuição de agrupamentos de genes de fumonisinas e giberelina é restrita a algumas espécies de *F. oxysporum* (219), e genes biossintéticos de fusarina, que estão amplamente distribuídas no *Fusarium* spp., estão ausentes em todos os isolados de *F. oxysporum* que foram examinados (162).

1.4.2 *A. thaliana* – *F. oxysporum*: interação hospedeiro – patógeno

Como exposto acima, a colonização de *A. thaliana* por *F. oxysporum* tem origem na região meristemática das raízes primárias ou secundárias (220) e induz uma reposta de defesa local ou sistêmica por esta. Além disso, dependendo de uma combinação específica entre a planta hospedeira e *F. oxysporum*, a resistência da planta ao fungo pode ser controlada por um gene (monogênica), por poucos genes (oligogênica) ou por múltiplos genes (multigênica) (182).

Os genes que codificam os receptores putativos ou receptores do tipo quinase (RTQ), At3g59700, At5g01540 e At5g60900 são constantemente induzidos, logo no primeiro dia pós-infecção. Em contraste, os genes At1g65790 e At4g13920 são induzidos posteriormente, no sexto dia após a infecção. A indução de um número de RTQs através da infecção por *F. oxysporum* pode ser devida à exigência de reconhecimento de múltiplos potenciadores liberados pelo fungo. Como a ativação de padrões de superfície celular (PSC), a regulação negativa da sinalização de PSC também é crucial. Ela previne a ativação excessiva ou prolongada das respostas imunes que poderão ser prejudiciais ao hospedeiro (221). Ligases de ubiquitina de domínio U-box, proteína U-box da planta (PUB)12 PUB12 e PUB13, são induzidas para este fim e recrutam um complexo de receptores, resultando em sua poliubiquitinação e endereçamento para o proteossoma para a degradação (222). A maior classe de proteínas PUB em *Arabidopsis*, cerca de 65%, é representada pelo domínio de repetição Armadillo (ARM) (223,224). Glickman e Ciechanover (2002) (225) demonstraram que ligação de moléculas de ubiquitina a uma proteína pode resultar em degradação proteossomal ou, alternativamente, regular outros processos, como transcrição e reparo de DNA. A regulação positiva dos genes U-box (PUB23) e de um par de genes contendo o domínio de repetição ARM observada em *Arabidopsis* infectada por *F. oxysporum*, sugere que as respostas imunes induzidas por parasitos também exigem uma sintonia de regulação fina e uma regulação negativa de sinais PSC. A indução de vários genes de resistência (R) de receptor *Toll* de interleucina 1 (TIR) foi também observada em *Arabidopsis*, através de infecção de *F. oxysporum* (221). At3g28220 (que codifica uma proteína contendo o domínio de MATH) e At1g52000 (que codifica uma lectina), ambos de uma família de proteínas que interagem com RPM1 4 (RIN4) e reguladores

negativos de imunidade de plantas (226), foram regulados positivamente em resposta a infecção por *F. oxysporum*.

Após o reconhecimento do patógeno pelas plantas, várias vias de transdução de sinal são ativadas. Os genes envolvidos em síntese de (i) ácido jasmônico (AJ), (ii) indol-3-ilmetil-glicosinolatos (I3G), (iii) biossíntese de camalexina (227–230), (iv) resistência a doenças, tais como PDF1.2 e PR4, e (v) de um modo mais geral os genes envolvidos nas vias de biossíntese de triptofano, I3G e camalexina são coordenadamente regulados positivamente em *Arabidopsis* infectada por *F. oxysporum* (229) assim que os genes do citocromo P450 e calose (221,226–229). Ativação da doença também ocorre para genes que codificam proteínas *dirigent-like* (At1g64160), proteínas *germin-like* (At1g18970), CAP (proteínas secretoras ricas em cisteína, Antígeno 5 e relacionada com patogênese 1) família de proteínas (At4g33720) e família de proteínas *wound-responsive* (At4g10270 e At4g33560) (221).

O papel das vias de sinalização mediadas pelo ácido salicílico (AS), o AJ e o etileno (ET) na resposta imune inata em *Arabidopsis* está bem estabelecido (231–233). Foi demonstrado que *F. oxysporum* é capaz de induzir resistência sistêmica adquirida (RSA), e proteínas relacionadas com a patogênese (PR) em *Arabidopsis*, indicando que a via AS desempenha um papel na resistência das plantas contra *F. oxysporum* (234). Além disso, o tratamento de folhas de plantas com SA antes da inoculação de *F. oxysporum* reduziu os sintomas da doença na planta (235). Ao analisar mutantes anormais nas vias de ET (*ein2-5*), AJ (*co1-1* e *jar1-1*) e AS (*NahG*, *sid2-1*, *eds5-1*, *npr1-1*, *pad4-1* e *eds1-1*), foi mostrado que as vias de AS, AJ e ET influenciam o resultado da doença de *F. oxysporum* em *Arabidopsis*. As vias de sinalização de SA, ET e JA interagem de forma positiva na ativação de resistência de *Arabidopsis* para *F. oxysporum*. A análise do transcriptoma feita por Zhu et al. (2013) (221) revelou que os genes que respondem ao estímulo do ET já estão ativados um dia após a infecção e que

os genes (i) envolvidos no metabolismo e sinalização de AJ, (ii) do metabolismo do triptofano que leva à biossíntese de auxina, (iii) que respondem ao AS, (iv) que são estimulados pelo ácido abscísico (ABA), e (vi) aqueles que estão mais geralmente envolvidos na regulação RSA foram ativados tardiamente, no sexto dia pós-infecção. Verificou-se que a expressão constitutiva de alguns reguladores de transcrição destas vias é suficiente para conferir maior resistência a *F. oxysporum* (236). Dos ~1400 fatores de transcrição (FTs) de acordo com Czechowski et al. (2004) (237), 12 foram regulados positivamente no primeiro dia pós-infecção e 10 destes FTs (MYB122; ANAC03, 42; AtbHLH038, 039; WRKY45, 51, 53, 63, 75) foram também regulados elevadamente no sexto dias após a infecção. Destes FTs constantemente induzidos, somente MYB122 e WRKY53 foram funcionalmente caracterizados (221).

A via do ABA também tem sido relacionada na resposta de defesa por meio da interação com outras vias e a regulação afinada da interferência entre estas vias parece determinar a saída de respostas de defesa das plantas a *F. oxysporum* (238). A sinalização do ABA, além de regular o desenvolvimento das plantas e resposta ao estresse abiótico, também desempenha um papel na regulação da imunidade inata (238–242). A meta-análise de genes de organismos patogênicos induzíveis em *Arabidopsis* revelou que um subconjunto significativo de genes regulados por ABA é ativado após infecção por patógeno (241). No entanto, na interação *A. thaliana* - *F. oxysporum*, o mutante *aba2-1*, que é deficiente na biossíntese de ABA, mostrou um aumento da resistência a *F. oxysporum*. Além disso, os mutantes *jin1-9* e *myc2*, que são deficientes no fator de transcrição MYC2, um regulador positivo de sinalização de ABA e um regulador negativo de resposta ao AJ, mostrou um aumento da resistência a *F. oxysporum* (231,243). Estes dados sugerem uma função negativa de ABA na resistência de *Arabidopsis* a *F. oxysporum*. No entanto, o mutante *ern1/irx1* (*enhanced resistance to necrotrophs 1*) que mostrou uma ativação constitutiva da via de ABA evidenciou um

aumento da resistência a *F. oxysporum* (244). Estes resultados contraditórios refletiram a complexidade do envolvimento da sinalização de ABA na resistência das plantas aos patógenos, em particular na interação *A. thaliana* - *F. oxysporum*.

A proteína G heterotrimérica de *Arabidopsis* foi descrita como sendo necessária para resistência a *F. oxysporum*. Proteínas G heterotriméricas são GTPases compostas de subunidades A, B e G que funcionam como mediadores na transdução de sinal dos diversos sinais externos em plantas, mamíferos e leveduras (245). Em plantas, proteínas G heterotriméricas também regulam várias vias de sinalização, tais como aquelas mediadas por auxina, giberelina e ABA (245–248). Além disso, a implicação das proteínas G heterotriméricas na produção e respostas de defesa às espécies de oxigênio reativo (ROS) foi confirmada em arroz (*Oryza sativa*) e *Arabidopsis* (246,248).

A resistência de *Arabidopsis* a *F. oxysporum* é positiva ou negativamente regulada por diferentes famílias de FTs. Um destes FTs é ATAF2, um membro da família de proteínas NAC, que é induzido pelo ferimento nas folhas, pelo tratamento com o AJ e AS, mas não com o ABA (249). A elevada expressão de ATAF2 em *Arabidopsis* aumenta a susceptibilidade a *F. oxysporum* e bloqueia a expressão de genes de defesa a *F. oxysporum* induzível, tais como PDF1.2 e PR1. ATAF2 tem sido proposto funcionando como um repressor de *F. oxysporum* induzível em respostas de defesa em *Arabidopsis* (250). As proteínas do fator de resposta ao etileno (FRE) pertencem a uma família de FTs composta de 122 membros em *Arabidopsis* (251). Vários FTs de FREs têm sido implicados diretamente na ativação ou inibição da resposta de defesa de *Arabidopsis* a *F. oxysporum*. Assim, a elevada expressão de FRE1, um integrador de ET e respostas ao JA (252), aumentam a resistência a *F. oxysporum* em *Arabidopsis*. A indução de FRE1 após desafio com o patógeno é bloqueada nos mutantes *coi1* e *ein3*, que são defeituosos nas vias de sinalização do AJ e do ET, respectivamente. Tal como acontece com FRE1, as plantas que sofrem uma elevada expressão do gene FRE2 eram

mais resistentes a *F. oxysporum* do que as plantas de tipo selvagem (253). Uma função semelhante na resistência a *F. oxysporum* foi descrita para FRE14, como a perda de função deste gene em mutantes, que mostrou um aumento da susceptibilidade a *F. oxysporum*. Em contraste, a inativação do FRE4 e AtMYC2 em *Arabidopsis* levou a um aumento da resistência a *F. oxysporum*, provavelmente através do aumento da resposta de defesa da planta de acordo com o AJ (231,253). Plantas transgênicas que possuem uma elevada expressão de FRE1 com expressão constitutiva da defensina antimicrobiana PDF1.2 e outras proteínas PR apresentaram resistência melhorada a *F. oxysporum*. Do mesmo modo, verificou-se que a elevada expressão de certas proteínas antimicrobianas, tais como, tioninas, é suficiente para conferir maior resistência a *F. oxysporum* em *Arabidopsis* (254) e tomate (255).

É importante notar que as respostas de defesa da planta estão interligadas com outros mecanismos de desenvolvimento, tais como fecho estomático (240), a senescência (256), floração (257), síntese da parede da célula (258), o metabolismo das giberelinas (259), a perda de cor (260) e ao estresse abiótico, o que sugere que, dependendo do tipo de estresse das plantas, algumas vias de sinalização poderiam ser dominante sobre as outras, adicionando ainda um outro nível de complexidade ao modelo de interações antagônicas apenas esboçado (231).

1.5 Modelagem por Análise de Balanço de Fluxo - FBA

A análise do fluxo de metabólitos através de uma rede metabólica pode ser feita através de uma abordagem matemática chamada análise de balanço de fluxo (*Flux Balance Analysis* – FBA) (261). O FBA tem como objetivo modelar a distribuição dos

fluxos metabólicos mais próximos dos naturais em um organismo alvo, a fim de prever a sua taxa de produção de biomassa, a qual pode ser assimilada a sua taxa de crescimento, ou a sua taxa de produção de um determinado metabólito (261–263). Também pode ser usado para analisar a rede metabólica em termos de estequiometria, termodinâmica e restrições de reação. Sujeito a essas restrições, as distribuições ótimas de fluxo metabólico podem ser calculadas usando programação linear. Calculando e examinando distribuições ótimas de fluxo sob várias condições, é possível gerar hipóteses quantitativas *in silico* que podem ser testadas experimentalmente (264).

No FBA, uma rede de genes conectados a reações catalisadas por enzimas é gerada através de expressões booleanas conhecidas como expressões de reação de proteína de gene (RPG). Uma matriz vinculando genes, proteínas e reações é usada para conectar os genes com a reação enzimática correspondente que os genes codificam. Cada nó da rede corresponde a um metabolito e as reações que os conectam entre si são representadas por arestas.

A hipótese central na qual FBA se baseia é que se o sistema é modelado em um estado estacionário, as concentrações de metabolitos estão em fluxo constante. Assim, as reações são escritas como equações matemáticas onde qualquer substrato que entra deve ser consumido, o que implica que a concentração interna de metabolitos dentro do modelo permanece constante ao longo do tempo. A hipótese de estado estacionário permite a redução do sistema metabólico a um conjunto de equações lineares, que são então resolvidas para encontrar uma distribuição de fluxo que satisfaça a condição de estado estacionário. O sistema de equações descreve as concentrações de metabolitos em uma rede como produto de uma matriz de coeficientes estequiométricos (a matriz

estequiométrica) e um vetor de fluxos não resolvidos. O lado direito do produto é um vetor de zeros que representam o sistema no estado estacionário. Uma abordagem conservadora é concordar que uma estratégia adequada de otimização biológica para uma rede metabólica de organismo completo é escolher o vetor de fluxo que maximiza o fluxo para a produção de biomassa.

Uma função objetiva, como a função objetiva da biomassa, é necessária para calcular um estado de rede realista e a distribuição de fluxo resultante (única ou não única) em uma reconstrução baseada em restrição, pois o espaço de soluções é geralmente muito grande para redes em escala genômica (265). Assim, pode-se escolher uma função de biomassa composta pelos metabólitos constituintes do organismo colocados na matriz estequiométrica e condicionar esta matriz estequiométrica para reduzir o espaço de soluções enquanto maximiza o valor de uma função que representa a conversão de precursores de biomassa em biomassa (266).

Do ponto de vista biológico, as restrições pertencem a dois subconjuntos que limitam (i) a absorção/excreção de nutrientes e (ii) o fluxo através de reações dentro do organismo. Essas restrições podem ser aplicadas para (i) a direção de uma reação ou (ii) a velocidade máxima de uma reação (267).

Embora a função objetiva mais utilizada no FBA seja a maximização da produção de biomassa (268), outros estudos utilizando diferentes funções objetivas já foram descritos. Exemplos de utilização de outras funções objetivas incluem a avaliação da capacidade de produção de diferentes metabólitos por um organismo celular, usando como objetivo a maximização da produção de alguns metabólitos primários ou

secundários (269,270), e o estudo da modelagem metabólica das mitocôndrias, utilizando como objetivo a maximização da função de produção de ATP (271).

Uma das limitações do FBA é a incapacidade de prever as concentrações de metabólitos por não usar parâmetros cinéticos, além de ser somente adequado para determinar os fluxos no estado estacionário. No entanto, em algumas apresentações, FBA não é capaz de explicar efeitos regulatórios, como por exemplo a regulação de expressão gênica ou ativação de enzimas por proteínas quinases, tornando as previsões do modelo imprecisas (261).

Com o grande número de organismos inteiramente sequenciados e a reconstrução de redes metabólicas em escala genômica é possível colocar interações patógeno-hospedeiro em uma perspectiva de rede metabólica a fim de prever possíveis mudanças metabólicas em todo o organismo resultantes das interações moleculares subjacentes (272).

Uma proposta seria a combinação da metodologia FBA para encontrar genes essenciais com testes de similaridade para determinar quais genes são específicos de um patógeno. Os genes que são bons alvos para interferência de RNA ou qualquer outro meio de inativação de alvos moleculares (inibidores de moléculas pequenas, peptídeos, haptâmeros, anticorpos) são aqueles que são (i) críticos para a produção de biomassa, no sentido FBA, e (ii) específicos para o genoma do patógeno como diagnosticado através de comparação de homologia de seqüência medida por *e-value* e/ou *score* (273).

2. OBJETIVOS

2.1 Objetivo Geral

Identificar uma estratégia de busca de alvos enzimáticos apropriada ao desenvolvimento de inibidores de patógenos em sistemas hospedeiro-parasitos.

2.2 Objetivos Específicos

- 1- Identificar as fontes de erro na busca de enzimas específicas baseada em EC nos pares *H. sapiens* e *L. major* usados como modelo.
- 2- Identificar enzimas específicas no sistema de relação parasitária entre *Arabidopsis thaliana* e *Fusarium oxysporum*; *Brassica rapa* e *F. oxysporum*; *Glycine max* e *F. oxysporum*; *Jatropha curcas* e *F. oxysporum*; *Ricinus communis* e *F. oxysporum*.
- 3- Identificar enzimas específicas de *F. oxysporum* cuja desativação tem maior potencial de desarticulação da rede metabólica deste fungo.
- 4- Identificar as consequências para *F. oxysporum* em termos de custo metabólico da inibição de alvos essenciais no modelo *A. thaliana* vs. *F. oxysporum*.

3. MATERIAL E MÉTODOS

3.1 Busca de Enzimas Específicas em *L. major* (Artigo publicado - Anexo 1)

3.1.1 Conhecimento utilizado

Devido a anotações imprecisas e fragmentadas depositadas em bancos de dados foi possível constatar que a ferramenta AnEnPi (37) é capaz de produzir diagnósticos falsos positivos. A partir deste fato, foi proposto um processo para depurar as fontes de erro.

A busca por falsos positivos foi realizada em uma lista de sequências enzimáticas obtida através da ferramenta AnEnPi, cujas funções foram anotadas como específicas para *L. major* em comparação com *H. sapiens*. Essa ferramenta classifica as enzimas em três categorias, enzimas homólogas, análogas e específicas, considerando as similaridades entre sequências com ≥ 100 aminoácidos e as anotações funcionais das enzimas nos dois organismos sendo comparados. Considerando todos os pares combinados de enzimas entre os dois organismos, uma enzima pode ser designada (i) homóloga a outra: quando a pontuação de similaridade e o *E-value* de seu alinhamento foram, respectivamente, ≥ 120 e perto ou abaixo de 10^{-4} (usando BLASTp); (ii) análoga a outra: quando a pontuação de similaridade de seu alinhamento for < 120 , mas as enzimas forem associadas ao mesmo EC (catalisando a mesma reação); (iii) funcionalmente específica: quando a enzima for homóloga a outra, conforme definido em "1", mas estiver associada a um EC diferente; e (iv) estritamente específica: quando não for homóloga e não compartilhar seu EC com nenhuma das outras enzimas. Em

outras palavras, ao classificar as enzimas como estritamente específicas para *L. major*, significa que as reações dessas enzimas são catalisadas em *L. major*, mas não em *H. sapiens*. Assim, uma enzima é considerada estritamente específica para *L. major* quando seu EC não é encontrado em *H. sapiens*.

Para este estudo, foi recuperada uma lista de genes proveniente do AnEnPi (versão de KEGG 58.1, 1 de junho de 2011) contendo 67 sequências de *L. major*, sendo que 42 dessas sequências estavam associadas a enzimas classificadas como estritamente específicas para *L. major* em comparação com *H. sapiens* por AnEnPi na base de seus respectivos ECs.

3.1.2 Processo de depuração de enzimas estritamente específicas

É possível distinguir várias etapas e componentes no processo de depuração de enzimas estritamente específicas. Primeiro, é necessário diagnosticar se as enzimas putativamente específicas de *L. major* possuem ou não homologia com o DNA humano e suas proteínas anotadas como enzimas. Em segundo lugar, uma enzima humana classificada como homóloga pode estar associada ao mesmo EC que *L. major* e pode ser diagnosticada como um falso positivo ou pode ser associada a um número de EC diferente e ser diagnosticada como funcionalmente específica. Em terceiro lugar, no caso de uma situação ambígua, é necessário a caracterização da estrutura bidimensional (2D) das regiões homólogas. Em quarto lugar, verificar mais profundamente as enzimas específicas e funcionalmente específicas através de análise tridimensional (3D) com Modelos ocultos de Markov (*Hidden Markov model*, HMM), *HHpred* e *HHsearch*, para detectar eventuais homologias remotas em referência às estruturas enzimáticas 3D humanas. Em quinto lugar, o conjunto de enzimas estritamente específicas de *L. major*

foi desafiado pela potencialidade de seus componentes como sendo alvos principais em referência à sua centralidade em caminhos metabólicos. Em sexto lugar, as enzimas específicas foram investigadas em outros parasitos humanos para avaliar sua relevância como alvos para o desenvolvimento de medicamentos.

O processo seguido para enriquecer as enzimas em proteínas estritamente específicas de *L. major* que são potencialmente adequadas para o desenvolvimento de fármacos é resumido na Figura 1. (i) A comparação de proteínas codificadoras de DNA de *L. major* com o DNA humano foi direta e a partir desta comparação foi possível obter as coordenadas onde as sequências homólogas poderiam ser encontradas no genoma humano (Figura 1A). A consistência das homologias detectadas pôde ser analisada através da realização de comparações múltiplas para determinar se as correspondências equivalentes foram obtidas de proteína para proteína e proteína para DNA usando como *queries* sequências do parasito para humano e humano para parasito (Figura 1B). (ii) A comparação da sequência de aminoácidos entre as enzimas de *L. major* e as proteínas humanas permitiu identificar rapidamente a maioria das proteínas humanas que mereciam atenção como potenciais enzimas homólogas (Figura 1C). (iii) A comparação das proteínas humanas identificadas em "ii" com o DNA cromossômico humano permitiu a delimitação de sua estrutura gênica (exons e introns). (iv) A comparação das proteínas de *L. major* que tiveram um *hit* homólogo com a sequência genômica humana sem uma proteína humana correspondente na lista proveniente do Ensembl (274) (http://www.ensembl.org/Homo_sapiens/Info/Index) indicou uma anotação de gene humano potencialmente ausente ou um possível pseudogene (Figura 1D). (v) A enzima *subject* de um par homólogo está associada a uma anotação inconsistente (Figura 1E) ou a um EC diferente em comparação a uma enzima *query* (Figura 1F). (vi) Os perfis 2D das sequências *query* e *subject* em pares homólogos

ambíguos foram comparados para discriminar casos de ortologia e paralogia distantes ou analogia (Figura 1G). (vii) As anotações ambíguas, como em "vi", também ocorreram quando uma enzima classificada como não homóloga ao genoma humano foi constatada compartilhando um mesmo EC com a de um homólogo em *L. major*, que é um caso de analogia intragenômica. Nesse caso, a enzima *L. major* não homóloga seria análoga à contraparte humana do homólogo, como poderia ocorrer quando o EC da contraparte humana estivesse incompleto ou ausente (Figura 1, 1H). (viii) As enzimas putativas estritamente e funcionalmente específicas de *L. major* para homologias remotas foram desafiadas com enzimas humanas utilizando as ferramentas HHpred e HHsearch (Figura 1I). (ix) As enzimas putativamente específicas de *L. major* foram desafiadas através de sua significância nas vias metabólicas do KEGG (Figura 1J). (x) Para avaliar a relevância das enzimas estritamente específicas de *L. major* para desenvolvimento de fármacos, a distribuição dessas enzimas foi analisada em parasitos humanos (Figura 1K).

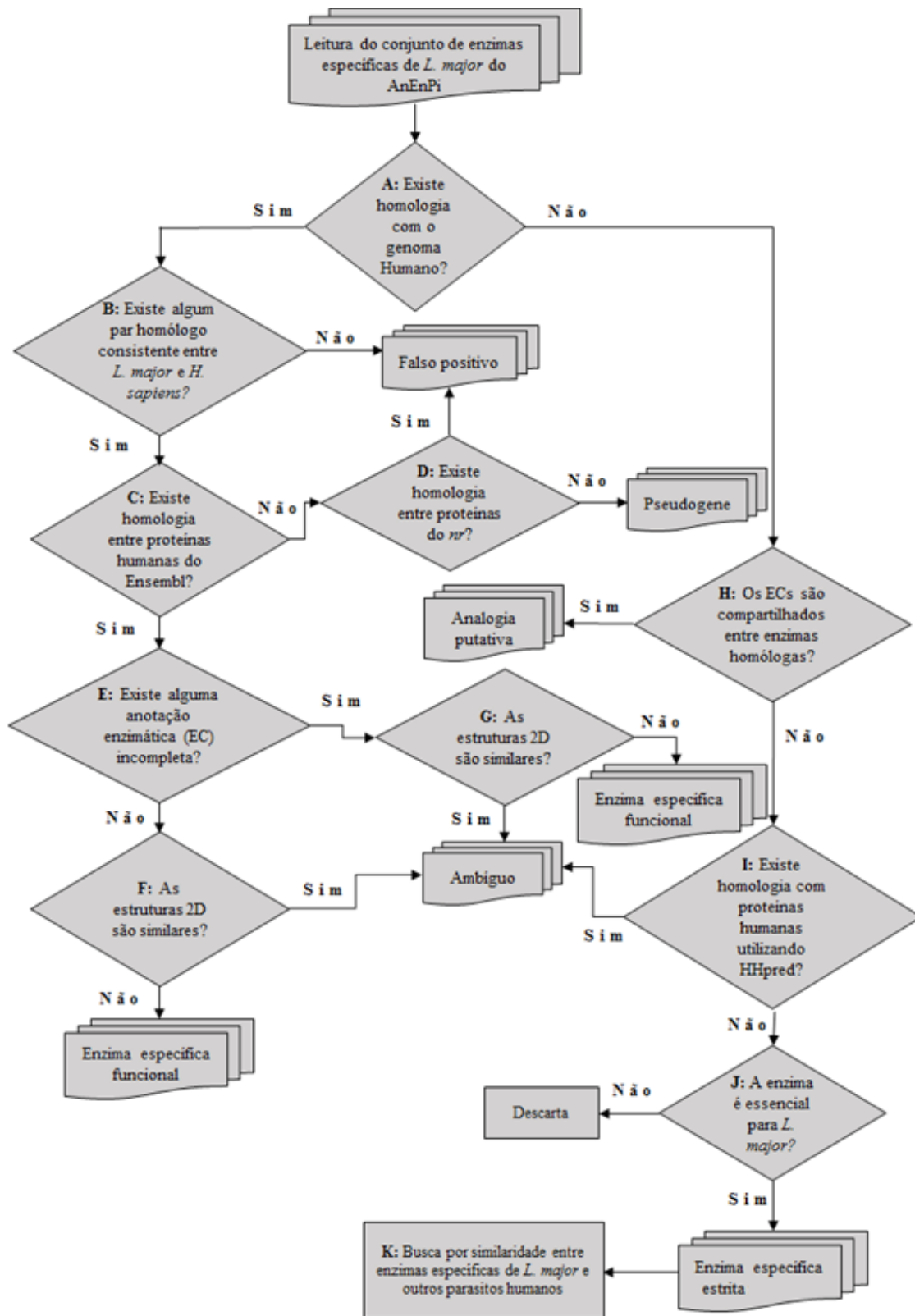


Figura 1: Fluxograma do processo de depuração de enzimas estritamente específicas.

3.1.3 Mapeamento de proteínas homólogas de *L. major* no genoma humano

Para determinar se a classificação de enzimas específicas poderia produzir falsos positivos, primeiramente foi realizada a busca por homologia entre as proteínas anotadas como específicas de *L. major* e a sequência de DNA do genoma humano como anotadas no Ensembl (<http://www.ensembl.org>) (Figura 1A). Para este fim, as sequências das enzimas supostamente específicas de *L. major* tal como documentadas no TriTrypDB (275) (versão 6.0-setembro de 2013) foram comparadas com a sequência do genoma humano (Ensembl, versão 74-novembro de 2013) utilizando tBLASTn. Para que o *hit* de um par de sequências fosse considerado um *hit* homólogo consistente, este teria que possuir um *E-value* de um alinhamento tBLASTn $\leq 10^{-4}$ e o valor de pontuação (*score*) ≥ 120 . Quando resultados consistentes foram obtidos, as coordenadas genômicas da saída do tBLASTn foram recuperadas e comparadas com as coordenadas genômicas das sequências de proteínas codificadas disponíveis do Ensembl (versão 74-novembro de 2013) para determinar se uma anotação de proteína poderia existir para a região genômica correspondente à região atingida durante o tBLASTn (Figura 2).

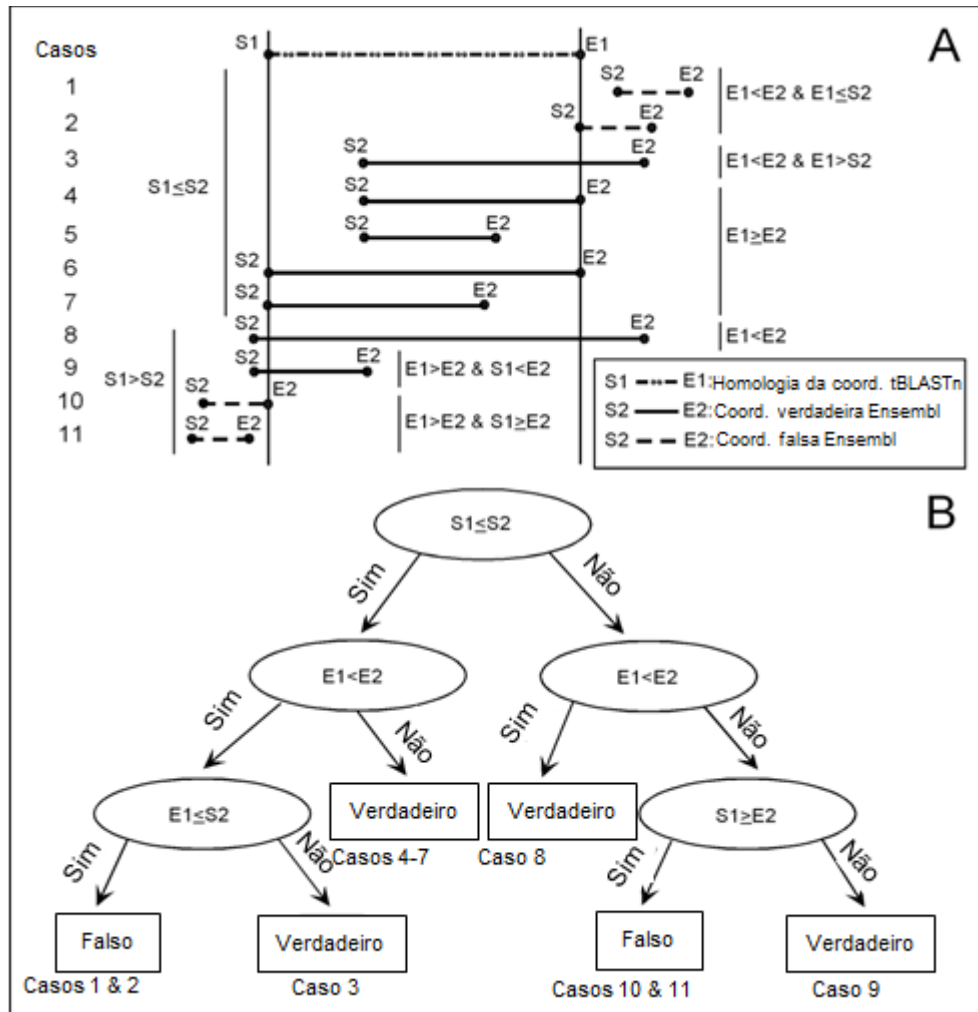


Figura 2: Algoritmo de busca homóloga para a classificação de falsos positivos de genes específicos de parasitas em seres humanos. A: As onze possibilidades das associações de proteína Ensembl (S2-E2) que se pode obter com uma região genômica humana que é homóloga (tBLASTn) a uma consulta de proteína parasita (S1-E1). S e E são para o início e o fim de uma homologia de tBLASTn ou coordenadas de genes humanos. A descrição booleana de cada associação entre um hit tBLASTn em uma região do genoma humano e as proteínas Ensembl anotadas nessa região é dada nos lados esquerdo e direito do painel A. '&' é usado aqui em seu sentido booleano, ou seja, um operador lógico AND. Os genes humanos para a proteínas Ensembl que são eventualmente compatíveis com a contraparte homóloga de um parasito (tBLASTn) são modelados por linhas finas. Os genes humanos para proteínas Ensembl na mesma região genômica que a contraparte homóloga de um parasito, mas que não

são compatíveis com ele, são representados por linhas tracejadas. B: A árvore de decisão para as associações VERDADEIRO e FALSO dos genes humanos para as proteínas Ensembl com uma região genômica humana que teve um hit homólogo (tBLASTn) com uma proteína de um parasito.

Como forma de automação desse processo, foram identificadas 11 possibilidades (Figura 2A), das quais sete são consideradas VERDADEIRAS e quatro são consideradas FALSAS no sentido booleano. As opções FALSA e VERDADEIRA podem ser facilmente diagnosticadas de acordo com a árvore de decisão mostrada na Figura 2B.

3.1.4 Consistência em pares homólogos de *L. major* vs. *H. sapiens* usando BLAST

A presença de introns nos genes de *H. sapiens* poderia complicar a comparação direta das sequências de genes de *L. major* com as do genoma humano, pois regiões homólogas, se presentes, poderiam ser interrompidas por um *splicing site* na sequência humana do gene homólogo. Para esclarecer este tipo de ambiguidade, foi realizada uma análise a fim de verificar se o *hit* genômico obtido pela comparação das proteínas de *L. major* com a sequência de DNA de cromossomos humanos (tBLASTn), conforme descrito acima, também corresponderia à região homóloga detectada na comparação por BLASTp (PSI-BLAST ou Position-Specific Iterative BLAST) da mesma *query* de *L. major* com a sequência da proteína humana (*subject*) correspondente a esse *hit* genômico. Assim, quando a anotação de uma proteína Ensembl para um *hit* homólogo entre uma proteína de *L. major* com o genoma humano existisse (Figura 1B, Figura 3A), a consistência com o modelo de gene, conforme encontrado pela comparação de

tBLASTn (Figura 3B) e das anotações *consensus coding sequence* (CCDS) pelo *National Center for Biotechnology Information* (NCBI) (Pruitt et al., 2009; Farrell et al., 2014) (Figura 3C) seria verificada e, caso consistente, seria atribuída à região genômica considerada onde sua homologia com a enzima *L. major* foi obtida.

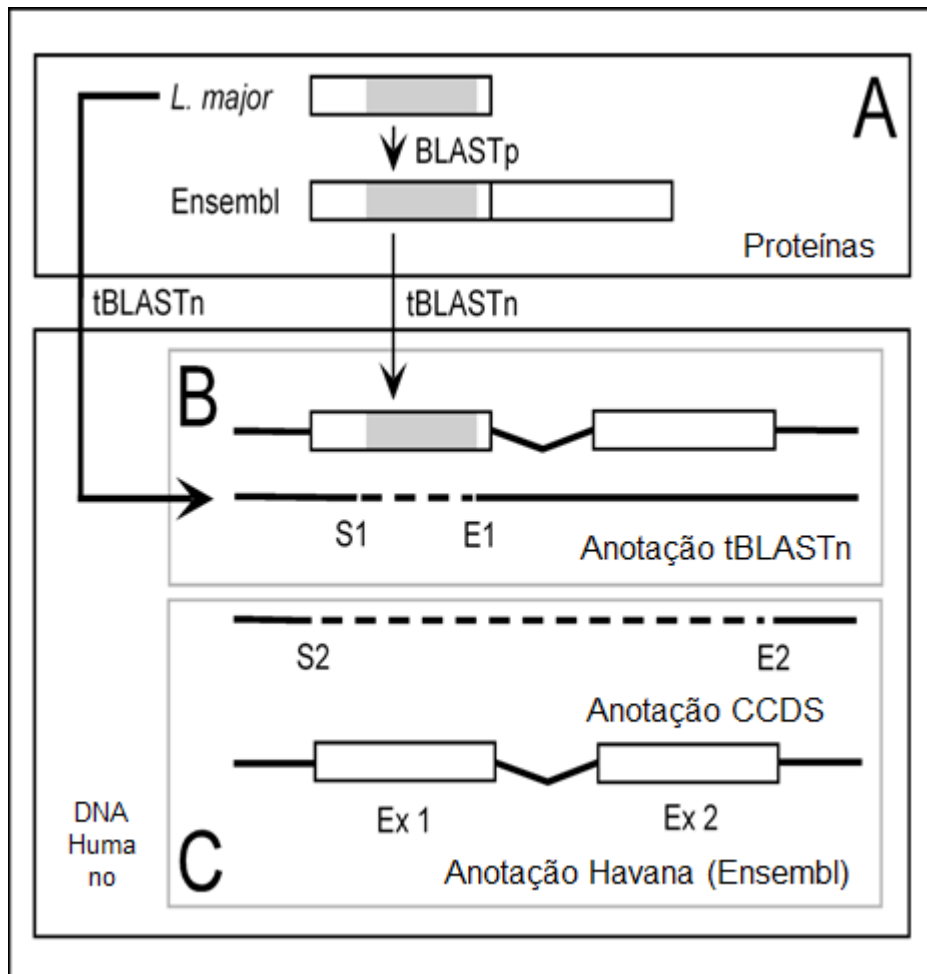


Figura 3: Esquema de pesquisa para falsos positivos de genes específicos de parasitos em seres humanos. O hit homólogo de *L. major* e proteínas humanas (A) é comparado à estrutura de genes humanos obtida pela pesquisa de tBLASTn (B) e à anotação de CCDS de Havana (C).

Para mapear uma região homóloga de uma proteína de *L. major* em seu gene humano marcado (Figuras 1A e 3), é necessário considerar várias linhas de evidência. (i) A homologia pode ser dividida em 2 regiões homólogas na sequência *subject* se um intron cair na sua região genômica correspondente ao domínio proteico da sequência *query* envolvida na homologia. Neste caso, as 2 homologias *subject* seriam contíguas na sequência *query* quanto à coordenada final da primeira região e à coordenada inicial da segunda região. (ii) A coordenada genômica da posição final de uma região homóloga menos a coordenada genômica de sua posição inicial mais 1 deve ser um múltiplo de 3 por causa da estrutura de códons das CCDSs. (iii) A região homóloga codifica uma proteína e, portanto, as coordenadas da sequência *subject* devem corresponder aos exons (Figuras 1B e 3B).

Após a comparação das proteínas de *L. major* com as proteínas humanas (BLASTp) no Ensembl (Figuras 1C e 3A), foi possível considerar várias etapas para automatizar o processo de mapeamento da região homóloga para éxons humanos (Figura 3). (i) Considerando a correspondência significativa desta comparação, foram recuperadas as coordenadas do gene humano de acordo com as sequências cromossômicas (Figura 3B). (ii) Para determinar o modelo de gene correspondente a uma proteína humana que tem um *hit* significativo com proteínas de *L. major*, foi realizada uma comparação (tBLASTn) entre a proteína humana proveniente do Ensembl com o trecho de sua sequência de DNA respectivo, extraíndo-o de suas coordenadas cromossômicas das posições gênicas correspondentes inicial e final conforme informado pela anotação do Ensembl (Figura 3C). (iii) As homologias detectadas pela comparação de uma proteína do Ensembl com o trecho correspondente do DNA humano, conforme descrito em "iii", forneceu as coordenadas para os éxons (*subject*) em relação a um modelo de gene. (iv) Para traduzir as coordenadas do éxon em "iii" em

coordenadas relativas ao contexto do cromossomo, foi adicionada as coordenadas do éxon à posição inicial do gene correspondente em coordenadas cromossômicas, conforme fornecido pelo Ensembl. (v) Para calcular o tamanho do intron, a coordenada cromossômica foi subtraída (em "iv") do início de um éxon a partir da coordenada cromossômica do final do éxon anterior para cada éxon do gene correspondente. (vi) Para calcular o tamanho do éxon, a coordenada cromossômica (em "iv") foi subtraída pelo fim de um éxon a partir da coordenada cromossômica pelo o seu início e somado mais 1; este passo foi realizado para todos os éxons de cada gene. Para tal, o tamanho de todos os éxons deveria ser um múltiplo de 3 (códon) para ser considerado relevante. (vii) A soma do tamanho do éxon dividido por 3 forneceu o tamanho da proteína, que foi inferido a partir do modelo derivado da sequência cromossômica. O tamanho da proteína pode então ser comparado com o tamanho da proteína de consulta do Ensembl para consistência de tamanho. (viii) Outro teste de consistência foi realizado determinando se as fitas ("+" ou "-") dos éxons, que foram deduzidas das homologias do *subject*, eram todas "+" ou todas "-". Éxons com alternância de fita "+ " e " - " para o mesmo gene com homologia com a mesma proteína *query* não fariam sentido, e esse procedimento permitiu diagnosticar casos de inconsistência. (ix) Foi possível confirmar os modelos de genes obtidos através da pesquisa tBLASTn das proteínas Ensembl com DNA humano, comparando as coordenadas do *hit* com as coordenadas de éxons disponíveis no banco de dados CCDS, conforme anotado pelo Ensembl. (x) Para mapear a região homóloga de uma sequência *query* de *L. major* com um éxon humano, as coordenadas cromossômicas (humanas) iniciais e finais (*query*) que correspondiam ao *hit* com as coordenadas inicial e final de cada éxon humano (obtidos em "iv"), respectivamente foram subtraídas. O éxon para o qual as duas subtrações obtiveram valores iguais a 0 (ou perto de 0, desde que a pequena extremidade saliente fosse um

múltiplo de 3 bases) foi considerado o que corresponde à região homóloga da sequência *query* de *L. major*.

As homologias entre uma proteína de *L. major* e uma proteína humana da lista do Ensembl foram consideradas como verdadeiras positivos quando (i) a proteína humana era legítima (todos os exons mapeados na mesma cadeia de DNA) e (ii) a região homóloga entre ambas as proteínas (*H. sapiens* e *L. major*) apresentavam uma combinação em seu quadro de codificação e as coordenadas da sequência genômica correspondiam ao *hit* da comparação (tBLASTn) das proteínas de *L. major* com o DNA cromossômico. Em outras palavras, para ser considerada significativa, uma homologia entre *L. major* e uma proteína humana deveria ser consistente com o modelo de gene dessa proteína humana (Figuras 1B e C e 3). Todas essas operações foram reunidas em uma única planilha do Excel (Tabela S1, acessível online: <http://journals.sagepub.com/doi/suppl/10.1177/1177932217712471> e <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5470852/>).

3.1.5 Hits genômicos sem contrapartes em proteínas do Ensembl

Nos casos em que um *hit* genômico obtido por tBLASTn não correspondeu à posição de qualquer proteína humana na lista do Ensembl (Figura 1D), este foi considerado como uma indicação de que o gene humano não recebeu uma designação durante o processo de anotação do Ensembl. A ausência de uma anotação foi abordada através da tradução do trecho homólogo do genoma humano (*subject*) em aminoácidos e usando a sequência de proteínas para uma pesquisa de homologia (BLASTp) com a seção *nr* do *GenBank* (versão 201.0-Abril de 2014). Quando uma proteína homóloga humana foi encontrada no *nr*, sua sequência completa foi comparada (tBLASTn) com

um trecho de DNA de 20 kb em torno do primeiro *hit* de *L. major* com o genoma humano para encontrar o modelo de gene (se houvesse) associado a esta proteína.

3.1.6 O caso particular da especificidade funcional

Considerando os casos de especificidade funcional enzimática entre as 2 proteínas homólogas de *L. major* e *H. sapiens* em que as proteínas foram anotadas como tendo diferenças na sua funcionalidade enzimática (com base no EC atribuído), as seguintes categorias foram caracterizadas (Figura 1E, F e H): (i) A incerteza com as anotações ocorreu quando pelo menos uma das funções enzimáticas em ambas as proteínas não pode ser confirmada (EC incompleto ou não disponível). Nesses casos, o nome e a função da proteína em UniProtKB / SwissProt (276) foram investigados a fim de verificar se a função descrita neste banco de dados corresponde à função descrita no banco de dados TriTrypDB. Se a função das enzimas que estavam sendo comparadas fosse um sinônimo em ambos os organismos, a proteína principal de *L. major* seria diagnosticada como um falso positivo para a especificidade funcional enzimática em relação à sua contraparte humana; caso contrário, a especificidade funcional seria atribuída à enzima de *L. major*. Esta posição conservadora foi tomada, pois caso os membros de um par homólogo tivessem a mesma função, a potência dos efeitos colaterais negativos sobre o hospedeiro pelo inibidor da enzima do parasito seria maior se as enzimas do hospedeiro e do parasito fossem ortólogas. (ii) A especificidade funcional enzimática resultante ocorreu nos casos em que a diferença na função enzimática das proteínas homólogas poderia ser confirmada. Quando ambos os genes do par são parálogos em vez de ortólogos, eles têm uma maior probabilidade de ter algum tipo de especificidade enzimática funcional (68). No entanto, um inibidor projetado

contra o alvo *L. major* ainda pode afetar a contraparte paróloga do hospedeiro a uma taxa mais baixa.

3.1.7 Mapeamento de domínio de proteínas e estrutura secundária

Para melhor caracterizar se os casos de anotações de enzimas ambíguas (Figura 1G) foram derivados de enzimas parálogas distantes ou análogas, foi realizada uma pesquisa para os núcleos catalíticos das enzimas mapeando a composição do domínio para as regiões conservadas dos pares homólogos de *H. sapiens* e *L. major* usando PROSITE (277), Pfam (278), e *The Conserved Domain Database* (279) para comparação da homologia utilizando os valores de identidade e *score*, $\sim 40\%$ e ≥ 120 , respectivamente. E a predição de seus perfis de estrutura secundária (2D) usando PRALINE (280).

3.1.8 Detecção de homologia remota com modelos ocultos de Markov

O BLAST é uma ferramenta de alta vazão para comparação de homologia de sequência, no entanto, as sequências podem ainda ser homólogas ao nível da estrutura terciária (3D) sem homologia reconhecível ao nível da sequência primária. A *twilight zone* é onde o nível de homologia (30-35%) possui a classificação ambígua (281). O modelo de Markov oculto (HMM) é adequado para detectar homologias distantes ($<30\%$) entre sequências, mas tem baixa vazão e, portanto, não é adequado para rastreamento de grandes bases de dados em pouco tempo e com poder computacional modesto. Como consequência, HMM é indicado para confirmar a especificidade de uma determinada enzima do parasito deduzida por busca de homologia com BLAST. Se um par homólogo

for detectado com HMM, comparando a enzima do parasito com os proteomas do seu hospedeiro, a relação de especificidade estabelecida anteriormente com BLAST deve ser descartada, senão ela pode ser mantida. Para esse fim, usamos as ferramentas do HH-suite (<https://en.wikipedia.org/wiki/HH-suite>) incluindo o *HMM-based lightning-fast iterative sequence search* (HHsearch) e *lightning-fast iterative protein sequence searching by HMM-HMM alignment* (HHblits) como descrito em Catharina et al. (2017) (282). Usamos a versão de HHsearch do servidor Mobylye (283) (<http://mobylye.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py>) assim que as versões de HHblits (<https://toolkit.tuebingen.mpg.de/#/tools/hhblits>) e HHpred (<http://toolkit.tuebingen.mpg.de/#/tools/hhpred>) disponíveis no servidor do instituto Max Planck em Tuebingen (284,285) (Figura 1I). Usamos as ferramentas (i) HHblits como método de geração de alinhamento (286) utilizando 75% como cobertura de sequência, 10^{-6} como *P-value* e 10 como número máximo de *hits*; (ii) HHsearch para busca de homólogo no banco *Protein Data Bank* (PDB; <http://www.rcsb.org/pdb/home/home.do>); e (iii) HHpred para predição de função e estrutura de proteínas com as opções padrão, ou seja, alinhamento local e classificação da similaridade da estrutura secundária.

3.1.9 Significância metabólica

A fim de esclarecer a relevância dos alvos enzimáticos putativos de *L. major* em termos da desorganização metabólica que se espera da inativação deles, seus respectivos EC foram buscados na base de dados da via KEGG (<http://www.genome.jp/kegg/pathway.html>). Em seguida, foi realizada uma busca por inserção de cada EC na via metabólica ao fim de diagnosticar se sua inativação

poderia afetar o parasito *L. major* sem ter consequências deletérias sobre o metabolismo humano e, conseqüentemente, se essas enzimas poderiam ser usadas como alvos para desenvolvimento de fármacos (Figura 1J).

3.1.10 Comparação de sequências de enzimas estritamente específicas de *L. major* com outros parasitos

Como uma enzima específica de *L. major*, em comparação com *H. sapiens*, também poderia ter pares homólogos em outros parasitos humanos (Figura 1K) e, eventualmente, pudesse servir como alvo para o desenvolvimento de medicamentos nesses organismos, homologias de sequência foram buscadas entre as enzimas estritamente específicas de *L. major* e as proteínas nos organismos listados na Tabela 1.

Tabela 1: Fonte de dados dos parasitas humanos importantes para comparação com enzimas estritamente específicas de *L. major*.

Organismo	GenBank		RefSeq	
	Assembly ID ¹	Atualização ID	Assembly ID ¹	Atualização ID
Fungi				
<i>C. neoformans</i>				
sorotipo JEC 21	091045.1	111678	091045.1	256138
sorotipo B- 3501 A	149385.1	172358	149385.1	172378
<i>H. capsulatum</i>				
sorotipo Nam1	149585.1	172628	149585.1	293588
<i>C. immitis</i>	149335.2	292488	149335.2	1741148
Helminths				
<i>S. mansoni</i>	237925.2	557858	237925.1	3549348
<i>B. malayi</i>	002995.2	449758	002995.3	451858
<i>L. loa</i>	183805.3	1163928	183805.2	4196238
Protozoan				
<i>T. gondii</i>	006565.2	751938	006565.2	3557728
<i>T. cruzi</i>	209065.1	271488	209065.1	271508
<i>T. brucei</i>	002445.1	522118	002445.1	2888
<i>L. major</i>	002725.2	283598	002725.2	402998
<i>L. donovani</i>	227135.2	405018	227135.1	405068
cepa BPK282A1				
<i>L. braziliensis</i>	002845.2	515698	002845.2	515718
cepa M2904				
<i>P. falciparum</i>				
sorotipo 3D7	002765.1	256198	002765.3	360518
sorotipo Dd2	149795.1	172858	NA ²	NA
sorotipo HB3	149665.2	2204658	NA	NA
<i>G. lamblia</i>	002435.1	278778	002435.1	284418
Bacteria				
<i>S. aureus</i> JH9	016805.1	17608	016805.1	37908
(MRSA/VRSA)				
<i>S. pyogenes</i> M1 GAS	<u>006785.2</u>	1009368	006785.2	1851728
<i>M. tuberculosis</i>	195955.2	538028	195955.2	538048
cepa H37Rv				

¹Assembly ID no formato XXX_000YYYYYY.Z onde X é o número identificador do banco como GCA e GCF que indicam a sequência proveniente do GenBank e RefSeq, respectivamente, Y indica o número de acesso e Z o número da versão. ²NA significa não associado.

Caso uma enzima destes organismos estivesse presente no KEGG (versão 81.0 - 1 de janeiro de 2017), sua anotação de EC seria comparada com a de *L. major*. Caso o EC não estivesse presente, não sendo possível uma comparação de anotações, seria necessária uma comparação de sequência utilizando o BLASTp. Para que as homologias com a sequência de *L. major* fossem significativas a região da homologia deveria ter uma cobertura de pelo menos 75% da sequência *query*, identidade $\geq 35\%$, uma *E-value* $\leq 10^{-4}$ e um *score* ≥ 120 (287).

3.1.11 Análise Filogenética

O efeito letal da inibição de REL1 foi mostrado em *Trypanosoma brucei* por Panigrahi et al. (288) que identificaram MP52 como a sequência associada a REL1 (289). A associação de sequências REL1 entre *T. brucei*, *Trypanosoma cruzi* e *L. major* foi claramente estabelecida pelo alinhamento (290) e permitiu atribuir inequivocamente a função REL1 a LmjF.01.0590. Por outro lado, as sequências completas do genoma foram produzidas para *Bodo saltans* e *Trypanoplasma borreli* que pertencem, respectivamente, a 2 ordens diferentes, isto é, Eubodonida e Parabodonida, em Metakinetoplastina (291). Essa relação permitiu testar a hipótese de se as sequências REL pertencentes ao EC:6.5.1.3 surgiu apenas dentro de *Trypanosomatidae* ou em outro nível taxonômico mais alto de Metakinetoplastina.

Das sequências completas do genoma de *B. saltans* (GenBank Assembly ID: GCA_001460835.1, ID da versão: 2707578) e *T. borreli* (<http://www.sanger.ac.uk/resources/downloads/protozoa/trypanoplasma-borreli.html>), foram encontradas sequências que codificam REL1 dentro da sequência NODE_50574 através de uma pesquisa tBLASTn de LmjF.01.0590 no arquivo Trypanoplasma_borreli_contigs_20120411.fa.gz.

As sequências REL1 de *T. brucei*, *T. cruzi*, *L. major*, *B. saltans* e *T. borreli* foram alinhadas utilizando a ferramenta ClustalW. As colunas correspondentes às lacunas foram eliminadas, e a melhor relação filogenética foi obtida usando o máximo de verossimilhança com MEGA7.57. O modelo de substituição foi Jones-Taylor-Thornton, a taxa entre os sites foi considerada uniforme e a inferência da árvore foi obtida por intercâmbio de vizinhos mais próximo (*Nearest Neighbor Interchange*).

3.2 Busca de Enzimas Específicas em *F. oxysporum* (Confirmação de Submissão de Artigo – Anexo 2)

3.2.1 Conhecimento utilizado

A base de dados do KEGG foi utilizada para obter anotações enzimáticas e suas correspondentes sequências de proteínas (versão 74.0, 1 de abril de 2015). Verificou-se que as reações enzimáticas eram específicas de *Fusarium graminearum* em comparação com *A. thaliana* quando os ECs da lista KEGG de *F. graminearum* não foram encontrados na lista de *A. thaliana*.

Na natureza, existe uma *formae speciale* de *F. oxysporum* que é patogênica para *A. thaliana*, no entanto, nem o genoma de *F. oxysporum* nem a sua anotação enzimática, EC, estavam disponíveis no KEGG quando o estudo foi realizado. Portanto, os dados de EC do fungo *F. graminearum* foram utilizados, uma vez que seus dados genômicos estavam disponíveis no KEGG. As duas listas de ECs de *F. graminearum* e *A. thaliana* foram retiradas do KEGG e comparadas para encontrar o subconjunto de ECs específicos de *F. graminearum*. Foram utilizadas anotações do KEGG para recuperar sequências de proteínas com os nomes de acessos associados ao ECs específicos de *F. graminearum*. Posteriormente, foi realizada uma comparação através de BLAST para verificar homologia entre as enzimas de *F. oxysporum* e as enzimas de *F. graminearum* caracterizadas como específicas em comparação com *A. thaliana*.

3.2.2 Pesquisa de enzimas ortólogas em *F. oxysporum*

Com o acesso do UniProtKB recuperamos as enzimas correspondentes aos ECs considerados específicos de *F. graminearum* e a partir destas sequências identificamos as enzimas ortólogas de *F. oxysporum* (cepa Fo5176) usando a ferramenta OrthoDB (292) (<http://www.orthodb.org/>). Quando uma função enzimática estava ausente, esta foi transferida de *F. graminearum* para *F. oxysporum* (cepa Fo5176) sob a condição de que fossem atingidos os limiares de identidade $\geq 35\%$, $E\text{-value} < 10^{-4}$, $score \geq 120$ e cobertura de alinhamento $\geq 75\%$ (BLASTp) (287).

3.2.3 *Hidden Markov model*

As sequências de proteína de *F. oxysporum* que são candidatas a serem específicas no contexto da relação hospedeiro-parasito com a planta *A. thaliana* foram submetidas separadamente à ferramenta HHblits *online* (<http://toolkit.tuebingen.mpg.de/hhblits>) sob os parâmetros padrão proposto para uma pesquisa de homologia usando o HMM como descrito em 3.1.8. As sequências de *F. oxysporum* que obtiveram um *hit* com qualquer sequência das quinze plantas utilizadas neste estudo foram consideradas falsas positivas e foram removidas da lista de proteínas específicas de *F. oxysporum* para o par hospedeiro-parasito considerado. Apenas as sequências que não produziram um *hit* com uma das quinze plantas na análise foram consideradas específicas do fungo.

3.2.4 *Depuração das enzimas específicas para casos falsos positivos*

Por homologia de sequência (tBLASTn), o subconjunto de enzimas específicas de *F. oxysporum*, como inferido de *F. graminearum*, foi comparado com o genoma de *A. thaliana* e de outras 14 espécies de plantas que são hospedeiras naturais de *F. oxysporum*: *B. rapa*, *Cajanus cajan*, *Carica papaya*, *G. max*, *Gossypium raimondii*, *J. curcas*, *Linum usitatissimum*, *Lotus japonicus*, *Medicago truncatula*, *Populus trichocarpa*, *Prunus pérsica*, *R. communis*, *Solanum lycopersicum*, *Theobroma cacao* (Tabela 2). Essas comparações adicionais permitiram descobrir a ocorrência de falsos positivos, mas também a conservação da especificidade de EC entre *F. oxysporum* em uma amostra significativa da sua gama de hospedeiros. No entanto, a depuração de dados para falsos positivos foi continuada apenas com *A. thaliana*, uma vez que esta é uma planta modelo utilizada para pesquisas científicas e com *B. rapa*, *G. max*, *R.*

communis e *J. curcas* devido às suas importâncias econômicas, em particular na produção de biodiesel. O procedimento descrito na seção 3.1 (282) foi aplicado de forma equitativa para as cinco espécies de plantas de acordo com os dados do genoma e proteína, conforme disponível nas bases de dados listadas na Tabela 3.

Tabela 2: Fontes dos dados de genomas completos e proteínas das plantas utilizadas para comparação de similaridade com *F. graminearum*.

Planta	Banco de dados	Versão e data da atualização	Número de Proteínas
<i>A. thaliana</i>	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/At_haliana/annotation/	V9.0; 13/12/2013	35386
<i>Brassica rapa</i>	ftp://brassicadb.org/Brassica_rapa/Bra_sc affold/	V1.2; 16/02/2013	41173
<i>Cajanus cajan</i>	http://gigadb.org/dataset/100028	V5.0; 2001/11/12	48680
<i>Carica papaya</i>	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Cp_apaya/annotation/	V9.0; 13/12/2012	27760
<i>Glycine max</i>	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Gmax/annotation/	V9.0; 13/12/2012	73320
<i>Gossypium raimondii</i>	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Gr_aimondii/annotation/	V9.0; 13/12/2012	77267
<i>Jatropha curcas</i>	ftp://ftp.kazusa.or.jp/pub/jatropha/	V4.5; 11/06/2012	57437
<i>Linum usitatissimum</i>	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Lu_sitatissimum/annotation/	V9.0; 13/12/2012	43484
<i>Lotus</i>	ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r2.5	V2.5;	38482

<i>japonicus</i>		21/10/2010	
<i>Medicago</i>	http://jcvi.org/medicago/display.php?	V4.0;	62317
<i>truncatula</i>	pageName=General&section=Download	31/07/2013	
<i>Populus</i>	http://dendrome.ucdavis.edu/ftp/Genome	V2; 26/08/2010	45033
<i>trichocarpa</i>	_Data/genome/Potr/v2/		
<i>Prunus</i>	ftp://ftp.bioinfo.wsu.edu/species/Prunus_	V1.0;	28702
<i>persica</i>	persica/Prunus_persica-genome.v1.0/genes/	10/04/2013	
<i>Ricinus</i>	ftp://ftp.jgi-	V9.0;	31221
<i>communis</i>	psf.org/pub/compgen/phytozome/v9.0/Rc	13/12/2012	
	ommunis/annotation/		
<i>Solanum</i>	ftp://ftp.solgenomics.net/tomato_genome/	V2.4;	34724
<i>lycopersicum</i>	annotation/ITAG2.4_release/	15/08/2014	
<i>Theobroma</i>	http://www.cacaogenomedb.org/Tcacao_g	V1.1;	46143
<i>cacao</i>	enome_v1.1#tripal_analysis-downloads-box	28/03/2012	

Tabela 3: Endereço eletrônico utilizados para obter o genoma, as sequências proteicas e genes de *A. thaliana*, *G. max*, *B. rapa*, *R. communis* e *J. curcas*.

	Genoma	Proteína	Gene
<i>A. thaliana</i>	https://www.Arabidopsis.org/download_files/Genes/TAIR10_genome_release/TAIR10_chromosome_files/TAIR10_chromosome_files	ftp://ftp.Arabidopsis.org/home/tair/Proteins/TAIR10_protein_lists/TAIR10_pep_20110103_representati	https://www.Arabidopsis.org/servlets/Search?action=new_search&type=gene

		ve_gene_model	
<i>G. max</i>	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Gmax/Gmax_189.fa	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Gmax/Gmax_189_protein.fa	plants.ensembl.org/Glycine_max/Info/Index
<i>B. rapa</i>	ftp://brassicadb.org/Brassica_rapa/Bra_Chromosome_V1.2/Brapa_sequence_v1.2.fa	ftp://brassicadb.org/Brassica_rapa/Bra_Chromosome_V1.2/Brassica_rapa_v1.2.pep	plants.ensembl.org/Brassica_rapa/Info/Index
<i>R. communis</i>	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Rcommunis/annotation/Rcommunis_119.fa	ftp://ftp.jgi-psf.org/pub/compngen/phytozome/v9.0/Rcommunis/annotation/Rcommunis_119_protein.fa	Castorbean.jcvi.org/downloads.php/TIGR_castorWG_S_release_0.1.genome.fsa
<i>J. curcas</i>	http://www.ncbi.nlm.nih.gov/assembly/GCF_000696525.1/GCF_000696525.1_JatCur_1.0_genomic.fna	/kegg/genes/organism/jcu/T03922.pep	http://www.ncbi.nlm.nih.gov/genome/

Resumidamente, as pesquisas de homologia (tBLASTn) entre as proteínas de *F. oxysporum* e genomas de plantas foram realizadas para verificar se a função de enzima fúngica foi codificada por uma região intergênica não anotada.

Em caso de *hit*, as coordenadas genômicas das posições inicial e final no cromossomo da planta foram recuperadas para identificar as proteínas codificadas pelas sequências nessas coordenadas. O arquivo de anotação de genes (Tabela 2), que contém informações sobre coordenadas genômicas e sequências das proteínas das plantas analisadas, foi usado para apoiar esta análise. Com isso, uma busca de similaridade com tBLASTn foi realizada entre as sequências protéicas das enzimas das plantas e suas respectivas regiões de DNA de plantas como codificadas pelos genes detectados nas coordenadas genômicas com mais ou menos de 10 kb. Para conferir a consistência de anotações, através do mapeamento de posição de exons e a identificação de seu quadro de leitura as homologias detectadas foram comparadas com as coordenadas de seus genes de codificação conforme documentado em seus arquivos de anotações *online*.

Em casos de um *hit* com uma região de DNA não anotada, a sequência de plantas homólogas no quadro de leitura (*query*) foi utilizada para uma busca por similaridade com as proteínas anotadas correspondentes na seção *nr* do GenBank através da interface *online* BLAST (www.blast.ncbi.nlm.nih.gov/Blast.cgi). Em seguida, o homólogo de planta que foi inferido foi confirmado por comparação com a região genômica a partir da qual foi extraída e, finalmente, as duas sequências de plantas e fungos foram comparadas por homologia e analogia como descrito anteriormente (seção 3.1; (282)).

3.2.5 Significância metabólica

A metodologia para obtenção da significância metabólica das enzimas foi realizada conforme o item 3.1.9 deste documento.

3.3 Validação das Enzimas Específicas por Análise de Balanço de Fluxo

– FBA

3.3.1 Análise de balanço de fluxo da reconstrução metabólica de *F.*

oxysporum

As reações metabólicas, que caracterizam a conversão de substratos em produtos, podem ser representadas matematicamente na forma de uma matriz estequiométrica S , onde $S_{m,n}$ representa o coeficiente estequiométrico do metabolito m ($m = 1, \dots, M$) na reação n ($n = 1, \dots, N$), que é negativo no caso de substratos e positivo de outra forma. Em tais modelos, o crescimento é representado por uma reação em que os substratos são precursores de biomassa com estequiometria igual à sua quantidade relativa na composição celular. No modelo de equilíbrio de fluxo do metabolismo celular, a célula é assumida em estado estacionário, onde a concentração de cada metabolito intracelular (diferente dos metabolitos que constituem a biomassa) permanece constante no tempo. Assim, as taxas de reação estacionárias (fluxos) que consomem e produzem um metabolito devem se equilibrar (equação 1),

$$\sum_{n=0}^N S_{mn} f_n = 0 \quad (1)$$

onde f_n denota o fluxo de reação n . O estudo do espaço de solução definido pela equação 1 juntamente com as restrições de fluxo para as taxas de absorção de substratos extracelulares constitui a base do FBA. Na maioria dos casos, há mais reações do que metabolitos e, conseqüentemente, soluções múltiplas para o problema podem-se encontrar. Uma forma de resolver esta ambigüidade é de identifica qual estratégia esta associada à produção máxima de biomassa entre as diferentes soluções já que um organismo sempre busca maximizar o seu crescimento. Se as relações das reações gene-

proteína forem fornecidas, relacionando genes a reações, pode-se preformar estudos de essencialidade genética que restrinjam o fluxo de reações associado a um gene específico para zero, o que é chamado de essencial se a remoção dele direciona a produção de biomassa para zero.

Como primeiro passo para a caracterização de alvos, encontramos o conjunto de genes essenciais relacionados à enzima na reconstrução metabólica de *F. oxysporum* (293). O modelo se encontra disponível no banco de dados Biomodels (<https://wwwdev.ebi.ac.uk/biomodels>) como MODEL1302010014 e compreende 2.346 reações, 2.170 metabolitos e 882 genes que fornecem relações de reações gene-proteína para 2.069 reações.

Finalmente, 6 genes representando 6 novas atividades enzimáticas foram inseridas de forma a atualizar a reconstrução metabólica de *F. oxysporum* com as enzimas específicas deste fungo em relação às cinco espécies de plantas, incluindo *A. thaliana*.

3.3.2 Detecção de homologia entre genes de *F. oxysporum* e *A. thaliana*

Consideramos como pares de genes homólogos os pares que apresentaram valores de $E\text{-value} \leq 10^{-4}$ utilizando a ferramenta BLAST para a identificação da homologia. Os pares de sequência que além de apresentar o valor de $E\text{-value} \leq 10^{-4}$ obtiveram uma pontuação de $score \geq 120$ foram classificados como homólogos e aqueles com $score < 120$ foram classificados como análogos (18).

3.3.3 Modelagem tridimensional das duas enzimas análogas consideradas críticas

As estruturas 3D de proteínas molde para modelagem *in silico* de proteínas por homologia foram obtidas no banco de estruturas do PDB usando o método HHsearch hospedado no Mobylye. Os modelos 3D de sequências *query* foram então gerados pelo procedimento HHalign-Kbest (294) que realiza alinhamentos HHM com base em sequências primárias e estruturas 3D de moldes de referência. Ao identificar um *hit* significativo, HHalign-Kbest gerou automaticamente de 20 a 50 modelos internamente e ranqueou os cinco melhores moldes classificados de acordo com os *Z-scores* calculados com Qmean4 (295). Entretanto, apenas o melhor modelo foi utilizado para a modelagem por homologia.

3.3.4 Análises físico-químicas

O cálculo de cargas eletrostáticas nos sítios ativos foi realizado com o servidor PDB2PQR utilizando a análise de Poisson-Boltzmann (296). O caráter de hidropatia foi determinado de acordo com a escala Kyte-Doolittle (297) com Chimera UCSF (298).

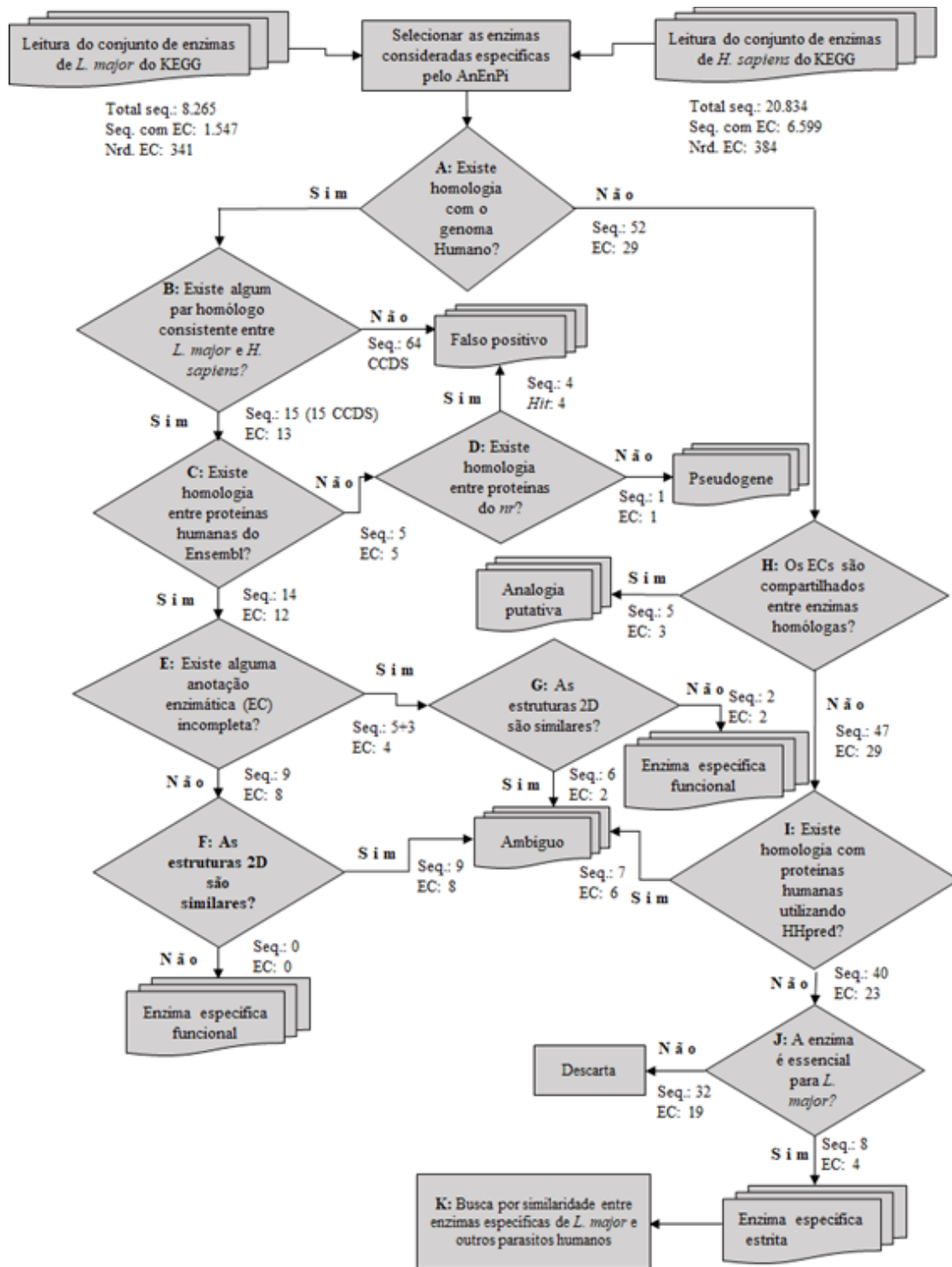
4. RESULTADOS

4.1 Busca de Enzimas Específicas em *L. major*

4.1.1 Metodologia de filtragem de enzimas estritamente específicas de *L. major* em comparação com *H. sapiens*

Comparando os conjuntos completos de (1) 8.265 sequências de *L. major*, sendo que 1.547 sequências possuíam ECs (341 ECs não redundantes) e (2) 20.834 sequências de *H. sapiens* com 6.599 sequências possuindo ECs (384 ECs não redundantes) provenientes do KEGG, a ferramenta AnEnPi obteve como resultado uma lista de 67 sequências de proteínas correspondendo a 42 atividades enzimáticas que eram consideradas específicas para *L. major* em comparação com os seres humanos.

A partir das 67 enzimas (42 ECs) obtidas pelo AnEnPi, 15 sequências (13 ECs) de *L. major* produziram 79 *hits* homólogos com o genoma humano em CCDSs, de acordo com a anotação feita pelo Ensembl, deixando 52 enzimas (29 ECs) de *L. major* sem nenhuma homologia com o genoma humano (Figura 4A). Ao comparar as coordenadas das estruturas genéticas (exons + introns) relatadas pelo Ensembl para as CCDSs com as coordenadas das homologias dadas pela pesquisa tBLASTn de sua sequência proteica com o genoma humano (exons putativos), foram identificadas muitas inconsistências nas anotações das CCDS do Ensembl (Figura 4B). Ao analisar os 79 *hits*, identificamos 64 CCDS não consistentes onde os exons putativos foram (1) associados à mesma proteína, mas apareceram em ambas as fitas “+” e “-” dos genes, o que é impossível dado o processo de tradução ou (2) simplesmente privado de uma associação proteica, impossibilitando qualquer inferência adicional (Tabela S1, *cf* acesso página 49).



Entre as 15 enzimas de *L. major*, 14 apresentaram homologia com as proteínas de *H. sapiens* de acordo com Ensembl e 1 enzima (LmjF.23.0270, pteridina redutase 1, EC 1.5.1.33) apresentou um *hit* com a própria sequência do genoma humano, mas não com as CCDSs anotadas pelo Ensembl (Tabela 4). A Tabela 4 mostra que, além de LmjF.23.0270, 4 outras enzimas de *L. major* também apresentaram *hits* com o genoma humano, mas que não foram anotadas como CCDS pelo Ensembl. No entanto, nestes 4 casos (LmjF.18.1510, LmjF.18.1520, LmjF.26.2280 e LmjF.27.2440), homologias também foram encontradas com CCDSs em outras coordenadas genômicas. No banco de dados do *nr*, foram encontradas sequências de proteínas homólogas para os *hits* genômicos humanos correspondentes a LmjF.23.0270, LmjF.18.1510, LmjF.18.1520 e LmjF.27.2440 que não foram anotados como CCDS pelo Ensembl. Foi possível recuperar um modelo de gene completo apenas para EAX0113, o que significou que a proteína hCG2039601 no cromossomo 18 na posição 19998852:19999097 bp não obteve uma anotação do Ensembl. Os modelos de genes para as 3 homologias nos cromossomos humanos 8 e 6 foram parciais, o que indica que provavelmente devem ser considerados como pseudogenes. Da mesma forma, LmjF.26.2280 (EC 3.5.5.1-nitrilase) correspondeu a um pseudogene, pois foi possível encontrar o códon de terminação ("*") na região homóloga humana (*subject*) no alinhamento tBLASTn (Figura 4D).

Tabela 4: Proteínas de *H. sapiens* recuperadas do *nr* utilizando o trecho de DNA do *subject* correspondente à região homóloga entre as proteínas de *L. major* e o genoma humano como *query*.

<i>L. major</i> id.	Chr # ¹	Coordenada	Proteínas Humanas Homólogas	
		<i>Hit</i> Genômico	<i>nr</i>	Acc. # ²
			Função	
LmjF.18.1510,	8	19098173	PMCA2a, parcial	NP_064524 ³
LmjF.18.1520		-19098421		
LmjF.18.1510,	18	19998852	hCG2039601	EAX0113 ⁴
LmjF.18.1520		-19999097		
LmjF.23.0270	6	99622932	3-hidroxibutirato	AAH01953 ³
		-99623354	desidrogenase,	
LmjF.27.2440	6	99622770	3-hidroxibutirato	ABG2424 ³
		-99623351	desidrogenase,	
LmjF.26.2280	1	78509203	tipo 2	NA
		-78509814	NA ⁵	

¹Chr # é para a identificação do cromossomo. ²Acc. # é para o número de acesso. ³Pseudogenes.

⁴Ausência de anotação. ⁵NA não associado.

Os acessos equivalentes do KEGG e os ECs das 17 proteínas humanas do Ensembl que foram homólogas às 14 enzimas de *L. major* (14 ECs e 3 repetições), apresentadas na Tabela 3, foram classificadas em 3 categorias (Figura 4C). (1) Nove enzimas foram anotadas com 8 ECs diferentes (1 repetição) em *L. major*, o que significa que elas foram associadas a diferentes reações enzimáticas em *L. major* e humanos, ou seja, elas podem ser consideradas específicas do ponto de vista funcional. (2) Três enzimas de *L. major* foram associadas a 2 ECs incompletas em humanos (1 repetição).

(3) Cinco (1 repetição na categoria "números EC diferentes") foram anotadas como enzimas em *L. major* (4 ECs), mas seus homólogos humanos não foram anotados como enzimas (indefinido) (ver Tabela 5). Vale ressaltar que, em todos os 17 casos, a região homóloga entre a proteína *L. major* e as sequências do genoma humano correspondeu precisamente a um exon exclusivo. Essas regiões conservadas, correspondentes aos domínios proteicos, nunca foram interrompidas por um intron no genoma humano.

Tabela 5: Anotações de números de EC de proteínas humanas homólogas com enzimas de L. major classificadas como específicas funcionais.

Acessos	EC #	Função Enzimática
Diferentes		
Números de EC		
LmjF.36.5960¹	3.1.4.46	Glicerofosfodiéster fosfodiesterase
Hsa:51573²/006007³	3.1.4.44	Glicerofosfoinositol glicerofosfodiesterase tRNA (5-metilaminometil-2-tiouridina)
LmjF.35.4250	2.1.1.64	-metiltransferase
Hsa:51805/132423	2.1.1.114	Poliprenil dihidroxibenzoato metiltransferase
LmjF.28.2910	1.4.1.4	Glutamato desidrogenase (NADP+)
Hsa:2747/182890	1.4.1.3	Glutamato desidrogenase [NAD(P)+]
LmjF.18.1510	3.6.3.6	H ⁺ -exportando ATPase
Hsa:478/105409	3.6.3.9	Na ⁺ /K ⁺ - trocando ATPase
LmjF.18.1520	3.6.3.6	H ⁺ - exportando ATPase
Hsa:478/105409	3.6.3.9	Na ⁺ /K ⁺ - trocando ATPase
LmjF.33.0680	1.3.1.71	Delta24(241)-esterol redutase
Hsa:7108/149809	1.3.1.70	Delta14-esterol redutase
LmjF.05.0350	1.8.1.12	Tripanotiona redutase
Hsa:7296/198431	1.8.1.9	Tioredoxina redutase (NADPH)
LmjF.09.0770	3.4.21.83	Oligopeptidase B
Hsa:5550/085377	3.4.21.26	Prolil oligopeptidase
LmjF.30.1550	4.2.1.70	Pseudouridilato sintase
Hsa:83480/110060	5.4.99.45	tRNA pseudouridina (38/39) sintase
Número de EC		
incompleto		
LmjF.26.2280	3.5.5.1	Nitrilase
Hsa:4817/158793	3.5.-.-	Nitrilase 1
LmjF.24.2030	1.1.1.100	3-oxoacil- [proteína transportadora de acil] redutase
Hsa:84869/145439	1.1.1.-	Oxidoreductase
LmjF.27.2440	1.1.1.100	3-oxoacil- [proteína transportadora de acil] redutase
Hsa:84869/145439	1.1.1.-	Oxidoreductase
Número de EC não		
associado		
LmjF.14.0350	4.2.3.1	Treonina sintase
Hsa: NA⁴/185875	-.-.-.-	
LmjF.30.1550	4.2.1.70	Pseudouridilato sintase
Hsa: NA/198331	-.-.-.-	
LmjF.30.0180	1.1.1.60	2-hidroxi-3-oxopropionato redutase
Hsa: NA/140632	-.-.-.-	
LmjF.18.1510,	3.6.3.6	H ⁺ - exportando ATPase

LmjF.18.1520

Hsa: NA/123191

-.-.-.-

¹ Números de acessos do KEGG para *L. major*. ² Números de acessos do KEGG para *H. sapiens*. ³ Números de acessos do Ensembl para *H. sapiens* de acordo com o formato ENSG00000xxxxxx. ⁴NA significa não associado.

Ao comparar as distribuições de α -hélice e de folha- β nos pares homólogos sob "número EC diferente" e "número EC incompleto" (Tabela 5), não foram encontradas diferença significativa nos perfis 2D, o que sugeriu que a classificação dessas sequências como funcionalmente específicas eram ambíguas. E somente em 5 sequências (4 ECs) sob "Número EC não atribuído" (Tabela 5) que diferenças significativas foram encontradas nos perfis 2D entre as duas regiões dos pares homólogos (Tabela 5, Figura 4F). No entanto, entre estas 5 enzimas, apenas 2 (LmjF.14.0350 e LmjF.30.0180) não estiveram envolvidas em homologias com as de "número de EC diferente". Essas 2 enzimas de *L. major* foram compostas por mais folhas- β do que α -hélices, o que não ocorreu em suas contrapartes humanas, onde a proporção entre folhas- β e α -hélices era aproximadamente a mesma.

A relação da análise de distribuição 2D entre os domínios conservados de *L. major* e humanos nos levou a considerar, finalmente, 2 sequências das 15 sequências (13 EC) que eram homólogas aos seres humanos como enzimas funcionalmente específicas em *L. major* em comparação com *H. sapiens*.

Considerando as enzimas putativamente específicas de *L. major* que compartilharam ECs com as sequências classificadas como enzimas funcionalmente específicas, 5 sequências (LmjF.18.0560, LmjF.19.1020, LmjF.26.0420, LmjF.36.1660 e LmjF.28.2100) compartilharam 3 ECs (Tabela 6). Como essas sequências são estritamente específicas, não se espera que elas tenham contrapartes humanas; no

entanto, o fato de que elas compartilham ECs com o grupo de enzimas que são homólogas ao genoma humano implica que elas poderiam ser análogas a estas sequências humanas. Essas 5 sequências foram desconsideradas, uma vez que o propósito era filtrar enzimas estritamente específicas.

Tabela 6: Enzimas putativas funcionais e estritamente específicas de *L. major* que dividem ECs.

<i>L. major</i>		<i>H. Sapiens</i> ¹		Classificação ⁴	
EC	Especificidade Funcional	Estrito	EC ²	Acesso. ³	EC
3.6.3.6	LmjF.18.1510	---	3.6.3.9 / NA ⁵	105409/123191	Homólogos
3.6.3.6	LmjF.18.1520	---	3.6.3.9 / NA	105409/123191	Homólogos
3.6.3.6	---	LmjF.18.0560	---	---	Análogos
4.2.1.70	LmjF.30.1550	---	5.4.99.45	110060	Homólogos
4.2.1.70	---	LmjF.19.1020	---	---	Análogos
4.2.1.70	---	LmjF.26.0420	---	---	Análogos
4.2.1.70	---	LmjF.36.1660	---	---	Análogos
3.1.4.46	LmjF.36.5960	---	3.1.4.46	006007	Homólogos
3.1.4.46	---	LmjF.28.2100	---	---	Análogos

¹ECs e números de acessos do Ensembl para sequências humanas homólogas com enzimas específicas putativas de *L. major*. ²Quando um EC humano não se encontrou disponível sob “número de EC não associado”, foi atribuída a classificação “Diferentes Números de EC” (Table 5) seguindo o padrão “Diferentes Números de EC”/Número de EC não associado”. ³O mesmo padrão foi atribuído em “²” para os números de acessos provenientes do Ensembl. ⁴Quando uma proteína do Ensembl homóloga a uma enzima específica putativa de *L. major* está disponível, a sua relação deve ser de homologia; quando esta enzima do Ensembl não está disponível, não existe homologia e o pareamento somente é através da função enzimática (EC), por exemplo, analogia. ⁵NA significa não associado.

4.1.2 Enzimas estritamente específicas de *L. major*

Através de uma verificação adicional de homologia remota com HHpred e HHsearch (Figura 4I), 7 sequências (LmjF.07.0270, LmjF.14.0180, LmjF.16.0530, LmjF.17.0140, LmjF.18.0200, LmjF.33.2540 e LmjF.36.3590) (6 ECs) foram consideradas homólogas distantes e foram filtradas a partir das enzimas putativamente específicas na Figura 4H (47 sequências, 29 ECs).

Uma análise de homologia das 40 sequências (23 ECs) restantes (Tabela 7) entre os *Trityps* e o *Plasmodium falciparum* mostrou que *L. major* (*query*) compartilhou: (1) 9 sequências com 24 sequências de *T. cruzi*, (2) 7 sequências com 18 sequências de *T. brucei* e (3) 2 sequências com 1 sequência de *P. falciparum* (Tabela 8).

Tabela 7: Enzimas específicas de *L. major* após filtragem com HMM.

Acesso	Função Enzimática	EC	Metabolismo
LmjF.07.0260	Homoserina desidrogenase	1.1.1.3*	Aminoácidos
LmjF.17.1360	L-gulonolactona oxidase	1.1.3.8*	Ascorbato
LmjF.34.0070	Ascorbato peroxidase, putativa	1.11.1.11	Ascorbato
LmjF.26.2610	Salicilato hidroxilase	1.14.13.1*	Degradação policíclica
LmjF.30.0610	Nitrato redutase, putativo S-metiltransferase de 5-	1.7.1.1*	Nitrogênio
LmjF.31.0010	metiltetra-hidropteril- triglictamato-homocisteína	2.1.1.14*	Aminoácidos
LmjF.36.2380, LmjF.36.2390	Esterol 24-c-metiltransferase, putativa	2.1.1.41	Esteróides
LmjF.09.1040	Fosfolipídico: diacilglicerol aciltransferase, putativa	2.3.1.158*	Glicolípideos
LmjF.30.3080	Homoserina quinase	2.7.1.39	Energético,

			aminoácidos
LmjF.17.1160	UDP-açúcar pirofosforilase	2.7.7.64	Carboidratos, Ascorbato
LmjF.11.1000	Piruvato fosfato diquinase, putativa	2.7.9.1	Fixação de carbono
LmjF.36.4640	Esterase de ácido ferúlico	3.1.1.73*	Carboidratos
LmjF.31.2300, LmjF.31.2310	Nuclease S1, putativa	3.1.30.1	Ácidos nucleicos
LmjF.04.0310, LmjF.04.0320, LmjF.23.0870, LmjF.23.0880, LmjF.35.0640	Beta-frutofuranosidase, putativa	3.2.1.26	Carboidratos
LmjF.29.2800 LmjF.10.0460, LmjF.10.0465, LmjF.10.0470, LmjF.10.0480, LmjF.31.2000, LmjF.28.0570	Purina nucleosidase, putativa GP63, leishmanolisina (protease)	3.2.2.1 3.4.24.36	Nucleótido
LmjF.34.3250	Indol-piruvato descarboxilase, putativa	4.1.1.74	Aminoácidos
LmjF.09.0360, LmjF.33.0470	DNA fotoliase	4.1.99.3	Nucleótido
LmjF.14.0460 LmjF.19.0985, LmjF.19.0995, LmjF.19.1005	Cistationina beta-liase 4-cumarato-COA ligase	4.4.1.8 6.2.1.12*	Aminoácidos
LmjF.26.0830	Aspartato - amônia ligase, putativo	6.3.1.1	Aminoácidos
LmjF.27.1870, LmjF.36.4300,	Tripanotona sintase, putativa	6.3.1.9	Glutaciona

LmjF.23.0460

LmjF.01.0590,

RNA ligase

6.5.1.3

Ácidos nucleicos

LmjF.20.1730

* EC sem nenhuma informação na literature de Tritryps para desenvolvimento de fármacos.

Tabela 8: Proteínas homólogas de *T. cruzi*, *T. brucei* e *P. falciparum* homólogas às enzimas específicas estritas de *L. major*.

<i>T. cruzi</i> vs. <i>L. major</i>		<i>T. brucei</i> vs. <i>L. major</i>		<i>P. falciparum</i> vs. <i>L. major</i>	
TcCLB.5037	LmjF.34.0070	Tb10.v4.0247	LmjF.36.2380	PFDG_021	LmjF.10.0460
45.30			LmjF.36.2390	71T0	LmjF.10.0465
TcCLB.5041	LmjF.36.2380	Tb427.01.3030	LmjF.20.1730		
91.10	LmjF.36.2390				
TcCLB.5041	LmjF.36.2380	Tb427.02.4370	LmjF.27.1870		
91.19	LmjF.36.2390				
TcCLB.5056	LmjF.36.2380	Tb427.10.6910	LmjF.36.2380		
83.10	LmjF.36.2390		LmjF.36.2390		
TcCLB.5061	LmjF.34.0070	Tb427.10.6950	LmjF.36.2380		
93.60			LmjF.36.2390		
TcCLB.5062	LmjF.11.1000	Tb427tmp.02.4150	LmjF.11.1000		
97.190					
TcCLB.5063	LmjF.20.1730	Tb427tmp.160.2970	LmjF.01.0590		
63.110					
TcCLB.5069	LmjF.20.1730	Tb927.1.3030	LmjF.20.1730		
75.9					
TcCLB.5081	LmjF.29.2800	Tb927.10.6910	LmjF.36.2380		
53.640			LmjF.36.2390		
TcCLB.5090	LmjF.09.0360	Tb927.10.6950	LmjF.36.2380		
27.10			LmjF.36.2390		
TcCLB.5101	LmjF.11.1000	Tb927.11.6280	LmjF.11.1000		
01.140					
TcCLB.5101	LmjF.01.0590	Tb927.2.4370	LmjF.27.1870		
55.20					
TcCLB.5101	LmjF.36.2380	Tb927.9.4360	LmjF.01.0590		
85.10	LmjF.36.2390				
TcCLB.5103	LmjF.29.2800	Tbg.972.2.2440	LmjF.27.1870		
15.10					
TcCLB.5103	LmjF.26.2610	Tbg972.1.1840	LmjF.20.1730		
41.4					
TcCLB.51158	LmjF.01.0590	Tbg972.10.8450	LmjF.36.2380		
5.20			LmjF.36.2390		

TCSYLVIO_	LmjF.11.1000	Tbg972.11.7080	LmjF.11.1000
001633			
TCSYLVIO_	LmjF.11.1000	Tbg972.9.2300	LmjF.01.0590
001634			
TCSYLVIO_	LmjF.01.0590		
003430			
TCSYLVIO_	LmjF.34.0070		
004399			
TCSYLVIO_	LmjF.29.2800		
004443			
TCSYLVIO_	LmjF.36.2380		
007533	LmjF.36.2390		
TCSYLVIO_	LmjF.20.1730		
008411			
TCSYLVIO_	LmjF.11.1000		
008551			

Das 40 enzimas estritamente específicas (23 ECs) (Tabela 7), 32 (19 ECs) foram desconsideradas, pois não pareciam estar relacionadas com enzimas essenciais em uma posição central nos mapas metabólicos, conforme documentado pelo KEGG (Figura 4J).

Apenas 8 enzimas específicas, correspondendo a 4 ECs (EC:2.1.1.41 com 2 sequências parálogas: LmjF.36.2380 e LmjF.36.2390; EC:2.7.9.1 com 1 sequência: LmjF.11.1000; EC:6.3.1.9 com 3 sequências parálogas: LmjF.23.0460, LmjF.27.1870 e LmjF.36.4300, e EC 6.5.1.3 com 2 sequências parálogas: LmjF.01.0590 e LmjF.20.1730), poderiam ser consideradas candidatas potenciais para o desenvolvimento de medicamentos. A relação completa de paralogia e analogia de genes e singularidade das 67 sequências obtidas pelo AnEnPi como enzimas supostamente específicas de *L. major* em comparação com humanos é dada na Tabela 9 e Tabela 10.

Tabela 9: Relação de paralogia ou analogia entre os grupos de proteínas de *L. major*.

Grupos ¹	<i>L. major</i> Acesso. ²	Número EC	Nome	Localização Gene		Relação	
				Crom. #	Início		Fim
			3-oxoacil- [proteína				
1	LmjF.24.2030	1.1.1.100	transportadora de	24	752837	753616	Analogia
	LmjF.27.2440		acil] redutase	27	1079832	1080713	Analogia
2	LmjF.36.2380	2.1.1.41	Esterol 24-C-	36	961263	962324	Paralogia
	LmjF.36.2390		metiltransferase	36	965198	966259	Paralogia
3	LmjF.31.2300	3.1.30.1	Nuclease S1	31	1138947	1140095	Paralogia
	LmjF.31.2310			31	1141911	1143047	Paralogia
4	LmjF.28.2100	3.1.4.46	Glicerofosfodiéster	28	819522	820571	Analogia
	LmjF.36.5960		fosfodiesterase	36	2296398	2297348	Analogia
5	LmjF.04.0310	3.2.1.26	Beta-	4	103123	105045	Paralogia
	LmjF.04.0320		frutofuranosidase	4	106566	108026	Paralogia
	LmjF.23.0870			23	430918	432576	Paralogia
	LmjF.23.0880			23	435934	437475	Paralogia
	LmjF.35.0640			35	302145	305417	Paralogia
6	LmjF.14.0180	3.4.17.19	Carboxipeptidase	14	46644	48158	Paralogia
	LmjF.33.2540			33	1138885	1140384	Paralogia
7	LmjF.10.0460	3.4.24.36	Leishmanolisina	10	213345	215153	Paralogia
	LmjF.10.0465			10	216396	218204	Paralogia
	LmjF.10.0470			10	219443	221377	Paralogia
	LmjF.10.0480			10	223401	225209	Paralogia
	LmjF.28.0570			28	201489	203189	Paralogia
	LmjF.31.2000			31	979311	981221	Paralogia
8	LmjF.18.0560	3.6.3.6	H ⁺ -exportando	18	228469	229707	Sem <i>hit</i> ³
	LmjF.18.1510		ATPase	18	679523	682447	Paralogia
	LmjF.18.1520			18	684689	687613	Paralogia
9	LmjF.09.0360	4.1.99.3	DNA fotoliase	9	145424	148228	Paralogia
	LmjF.33.0470			33	207188	208813	Paralogia
10	LmjF.19.1020	4.2.1.70	Pseudouridilato	19	420290	421867	Analogia

		sintase					
	LmjF.26.0420			26	114341	116431	Analogia
	LmjF.36.1660			36	652829	654913	Analogia
	LmjF.30.1550			30	548003	549499	Analogia
		4-cumarato-COA					
11	LmjF.19.0985	6.2.1.12		19	398614	400455	Paralogia
		ligase					
	LmjF.19.0995			19	403723	405531	Paralogia
	LmjF.19.1005			19	410749	412560	Paralogia
12	LmjF.23.0460	6.3.1.9	Tripanotiona sintase	23	174520	175317	Paralogia
	LmjF.27.1870			27	806164	808122	Paralogia
	LmjF.36.4300			36	1657674	1659923	Paralogia
13	LmjF.01.0590	6.5.1.3	RNA ligase	1	163208	164482	Paralogia
	LmjF.20.1730			20	715134	716390	Paralogia

¹ Cltrs para os grupos.

² Acc. para os números de acessos das proteínas de *L. major*.

³ Sem *hit* para as enzimas agrupadas no mesmo grupo.

Tabela 10: Enzimas específicas cujos genes são únicos no genoma de *L. major*.

Acesso <i>L. major</i>	Número EC	Localização Gene		
		Cromo.	Início	Fim
		#		
LmjF.07.0260	1.1.1.3	7	104668	103568
LmjF.17.1360	1.1.3.8	17	637740	639248
LmjF.34.0070	1.11.1.11	34	22865	23776
LmjF.26.2610	1.14.13.1	26	1052042	1053406
LmjF.30.0610	1.7.1.1	30	208677	207067
LmjF.16.0530	1.3.3.1	16	192940	191978
LmjF.31.0010	2.1.1.14	31	5558	3246
LmjF.09.1040	2.3.1.158	9	428023	425966
LmjF.36.3590	2.5.1.47	36	1394413	1393412
LmjF.30.3080	2.7.1.39	30	1161470	1160445
LmjF.17.1160	2.7.7.64	17	567197	569089
LmjF.11.1000	2.7.9.1	11	407502	410246
LmjF.36.4640	3.1.1.73	36	1794330	1796072
LmjF.29.2800	3.2.2.1	29	1184172	1185173

LmjF.17.0140	3.4.11.14	17	108084	106813
LmjF.07.0270	3.5.1.16	7	108286	107093
LmjF.34.3250	4.1.1.74	34	1443491	1445143
LmjF.14.0460	4.4.1.8	14	152236	153615
LmjF.18.0200	5.4.99.9	18	100496	99021
LmjF.26.0830	6.3.1.1	26	243130	242069
LmjF.30.0180	1.1.1.60	30	54045	53146
LmjF.33.0680	1.3.1.71	33	287058	288383
LmjF.28.2910	1.4.1.4	28	1109443	1108085
LmjF.23.0270	1.5.1.33	23	101918	102784
LmjF.05.0350	1.8.1.12	5	105666	107141
LmjF.35.4250	2.1.1.64	35	1715055	1716020
LmjF.09.0770	3.4.21.83	9	321130	318935
LmjF.26.2280	3.5.5.1	26	918955	919794
LmjF.14.0350	4.2.3.1	14	104841	106871

4.1.3 Relacionamento evolutivo de enzimas estritamente específicas em *L. major* com outros parasitos humanos

Para entender melhor as consequências do relacionamento evolutivo de enzimas estritamente específicas para o desenvolvimento de fármacos (Figura 4K), foi realizada uma busca por proteínas homólogas em outros parasitos humanos (Tabela 6). O EC:2.1.1.41 (esterol 24-C-metiltransferase) parece ser conservado entre os fungos e *L. major*, o que é esperado devido ao seu uso comum na via ergosterol. Curiosamente, as homologias para o EC:2.7.9.1 (piruvato fosfato diquinase) são apenas compartilhadas entre os protozoários; a enzima de *Mycobacterium tuberculosis* deve ser considerada análoga, pois não compartilha homologias com as sequências de *L. major*. Como pode ser observado na Tabela 6, o EC:6.3.1.9 (tripanotiona sintetase) é estritamente específico para *Trityps*, o que confirma as pesquisas anteriores. À primeira vista, o EC:6.5.1.3 (RNA ligase) é uma função que parece relativamente bem conservada entre parasitos humanos. No entanto, REL1 (LmjF.01.0590) e REL2 (LmjF.20.1730) não são

homólogas à ligase de transferência de RNA (tRNA) de *P. falciparum* e *Toxoplasma gondii* (Apicomplexa), mesmo que tenham sido anotadas com a mesma função enzimática (EC:6.5.1.3). Assim, os RELs de *Trypanosomatidae* e RtcBs de Apicomplexa devem ser considerados análogos. A árvore filogenética na Figura 5 mostra que REL1 é conservada em Metakinetoplastina até *B. saltans* e *T. borreli* que são representantes desta subclasse, juntamente com *Trypanosomatidae*. O REL2 não foi levado em consideração, pois foi provado que não é essencial em *T. brucei* (288).

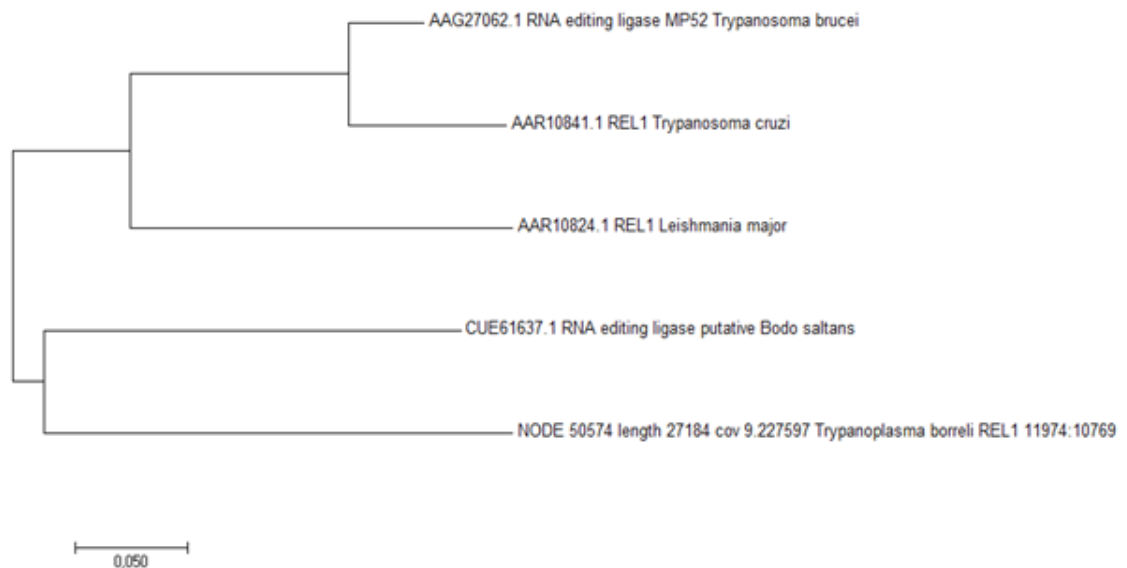


Figura 5: Árvore filogenética de REL1 em Metakinetoplastina obtida por máxima de verosimilhança. Escala: 0,05 substituições por sítio.

4.2 Busca de Enzimas Específicas em *F. oxysporum*

4.2.1 Comparação de EC e recuperação da sequência de proteínas

Conforme descrito na metodologia, consideramos na primeira aproximação que *F. graminearum* poderia ser um representante adequado de *F. oxysporum*. Assim, ao comparar os hospedeiros vegetais com *F. graminearum* como representante de *F. oxysporum*, descobrimos que, dos 7.008 ECs não redundantes do KEGG, 879 foram associadas a *F. graminearum* e 1.044 a *A. thaliana*, respectivamente.

Ao comparar ambas as listas, identificamos 247 EC que eram específicas de *F. graminearum*, das quais recuperamos 318 sequências de proteínas. A comparação através de BLASTp destas 318 sequências de proteínas de *F. graminearum* com as proteínas anotadas (proteomas) das quinze espécies de plantas deste estudo: *A. thaliana* (n = 35.386), *B. rapa* (n = 41.173), *C. cajan* (n = 48.680), *C. papaya* (n = 27.760), *G. max* (n = 73.320), *G. raimondii* (n = 77.267), *J. curcas* (n = 57.437), *L. usitatissimum* (n = 43.484), *L. japonicus* (n = 38.482), *M. truncatula* (n = 62.317), *P. trichocarpa* (n = 45.033), *P. persica* (n = 28.702), *R. communis* (n = 31.221), *S. lycopersicum* (n = 34.724), *T. cacao* (n = 46.143) resultou em uma diminuição do número de enzimas que poderiam ser consideradas como específica de *F. graminearum* para 83, o que sugere que os ECs que são específicos de *F. graminearum* comparadas com essas espécies de plantas pode variar de caso para caso e que o número de ECs que são específicos de *F. graminearum* diminui de acordo com o aumento de tamanho da amostra de hospedeiros da planta considerados na comparação BLASTp.

4.2.2 Busca por enzimas específicas em *F. oxysporum*

Em uma primeira tentativa, verificamos a homologia remota das 83 proteínas específicas de *F. graminearum* com o proteoma de *A. thaliana* usando HHblits e encontramos hits para 50 deles. Portanto, essas enzimas de 50 *F. graminearum* foram descartadas de análises adicionais. O conjunto restante continha apenas 33 enzimas aparentemente específicas para *F. graminearum*.

Neste ponto, poderia ser realizada uma procura por enzimas específicas de *F. oxysporum* em vez de *F. graminearum*. Assim, a partir das 33 enzimas específicas de *F. graminearum* com acessos do KEGG, foram recuperados os 33 números de acesso correspondentes do UniProtKB. Uma enzima, FG10182.1 (EC:4.3.2.3) foi atualizada para FG13713 na versão 83.2 (1 de setembro de 2017) do KEGG. Ao pesquisar os ortólogos destas 33 enzimas específicas de *F. graminearum* na cepa Fo5176 de *F. oxysporum* com a ferramenta OrthDB, foram recuperadas 30 proteínas (Tabela 11), o que significa que 3 enzimas FG03537.1 (EC:4.2.3.6), FG10343. 1 (EC:3.4.14.9) e FG11037.1 (EC:.2.1.151) de *F. graminearum* não possuem pares ortólogos no genoma de *F. oxysporum*.

A taxa de similaridade entre as 30 enzimas específicas de *F. oxysporum* e *F. graminearum* foi bastante alta, já que descobrimos que a região de similaridade cobria, em média, 97% do tamanho da *query* para os pares homólogos entre os dois fungos. Apenas um par homólogo tinha uma região de similaridade cobrindo 67% da sequência da *query*, mas envolvia uma região homóloga de 368 aminoácidos com uma taxa de identidade de 77%. As outras regiões de similaridade entre pares homólogos são responsáveis por pelo menos 80% do tamanho da *query*. A identidade média dos pares homólogos entre as 30 sequências específicas foi de 83,7%, com apenas duas

sequências com um nível de identidade inferior a 70%, isto é, 57% com uma cobertura de 100% e 66% com uma cobertura de 91%. Estes resultados mostram que as funções podiam se transferidas de *F. graminearum* para *F. oxysporum* com um excelente nível de confiabilidade.

Embora apenas 11 proteínas de *F. oxysporum* de UniProtKB tenham uma atividade enzimática atribuída, foi possível confirmar com sucesso todos os casos de ortologia por BLASTp e, portanto, foi possível transferir o EC correspondente do KEGG quando necessário. Assim, no passo seguinte, foram removidas da análise as enzimas FG03537.1, FG10343.1 e FG11037.1 e foi considerado o conjunto composto de 30 enzimas descritas acima como representativas do conjunto de enzimas específicas de *F. oxysporum*.

4.2.3 Depuração de anotação de homólogos de enzimas hospedeiras

As 30 enzimas consideradas específicas para *F. oxysporum* com base em sua comparação proteômica com as de seus hospedeiros vegetais foram associadas a 25 atividades enzimáticas (EC:1.13.11.1, 1.13.11.52, 1.13.11.6, 1.8.4.14, 2.1.1.17, 2.1.1.286, 2.1.1.312, 2.3.1.7, 2.4.1.16, 2.7.7.53, 3.1.3.84, 3.1.4.12, 3.2.1.101, 3.2.1.39, 3.2.1.40, 3.2.1.8, 3.4.13.19, 3.4.14.4, 3.6.1.59, 4.1.1.6, 4.2.1.79, 4.2.1.94, 4.3.2.3, 4.4.1.17, 5.5.1.5) (Tabela 3). Ao pesquisar essas 30 enzimas para similaridade (tBLASTn) com os genomas de *A. thaliana*, *B. rapa*, *G. max*, *R. communis* e *J. curcas*, foram encontradas apenas duas enzimas (F9G623 e F9FNN9) que não produziram *hit* com qualquer uma das cinco espécies de plantas e, portanto, poderiam ser consideradas dois casos de verdadeiros positivos.

Em um caso, a comparação de F9G156 produziu um *hit* com a região genicas de *A. thaliana* correspondente a AT5G65060.1 e AT5G65050.3. Quando essas regiões homólogas foram comparadas com *nr*, elas produziram *hits* significativos para o fator de transcrição MADS-box de *F. oxysporum* (FOSC 3-a), o que sugere que F9G156 não é específico para *F. oxysporum* quando comparado a *A. thaliana*.

O outro caso foi relativo a F9FMM7 que deu um *hit* com o genoma de *J. curcas* correspondente à proteína jcu: 105635122. Ao comparar (BLASTp) a região homóloga com *nr*, foram obtidos *hits* significativos entre F9FMM7 e espécies do gênero *Fusarium*, o que sugere que essa enzima fúngica não é específica do parasito em *J. curcas*.

4.2.4 Significância Metabólica

Os ECs atribuídos a cada enzima de *F. oxysporum* que foram mapeados com sucesso em caminhos metabólicos do KEGG (*KEGG PATHWAY*) foram inseridos em 14 mapas metabólicos diferentes (Tabela 11). Essas enzimas foram distribuídas em ECs com uma única via de execução e ECs com rotas alternativas de execução, ou seja, diferentes rotas metabólicas para um mesmo produto. As enzimas distribuídas em EC com uma única via de execução que foram denotadas como enzimas primárias foram: 2.4.1.16, 3.6.1.59, 2.3.1.7, 1.13.11.6, 4.3.2.3, 4.4.1.17, 3.1.4.12, enquanto as ECs com rotas alternativas de execução que foram designadas enzimas secundárias foram: 4.2.1.79, 1.13.11.52, 2.1.1.17, 1.8.4.14, 2.7.7.53.

Tabela 11: Enzimas consideradas específicas a *F. graminearum* (*FGR*) e *F. oxysporum* cepa *Fo5176* (*FOX*) em relação a *A. thaliana*, *G. max*, *B. rapa*, *R. communis*, *J. curcas*.

#	FGR	FGR acesso ¹	FOX acesso	EC	Nome	Via
	acesso	(UniProtKB)	(UniProtKB)			metabólica

(KEGG))			(KEGG)
1	FG02259.1	IIREZ9	F9FYW8	2.7.7.53	ATP adenililtransferase Fgr00230
2	FG10182.1 ²	IISA32	F9F7R0	4.3.2.3	Ureidoglicolato liase
3	FG10062.1	A0A098DA90	F9F7C4	1.8.4.14	L-metionina (R) -S- óxido redutase Fgr00270
4	FG04828.1	IIRLM2	F9FVQ5	1.13.11.52	Indoleamina 2,3- dioxigenase Fgr00380
5	FG07058.1	IIRSE0	F9G623	1.13.11.6	Hidroxiantranilato 3,5 dioxigenase
6	FG01272.1	A0A0E0RQ52	F9F1B7	2.4.1.16	Quitina sintase Fgr00520
7	FG10116.1	IIS097	F9FUJ3	2.4.1.16	Quitina sintase
8	FG05066.1	A0A0E0SIE3	F9FNN9	2.1.1.17	Fosfatidiletanolamina N-metiltransferase Fgr00564
9	FG01057.1	IIRBW6	F9FM77	3.1.4.12	Esfiromielina fosfodiesterase 2 Fgr00600
10	FG08851.1	IIRX10	F9G156	4.2.1.79	2-Metilcitrato desidratase Fgr00640
11	FG08386.1	IIRVU1	F9FX57	4.4.1.17	Holocitocromo-c sintase Fgr00860
12	FG03537.1	A0A1I9FCV5	NA	4.2.3.6	Tricodieno sintase Fgr00909
13	FG09810.1	IIRZG9	F9FE93	3.6.1.59	m7GpppX difosfatase Fgr03018
14	FG00840.1	V6QVV9	F9FPR6	2.3.1.7	Carnitina acetiltransferase Fgr04146
15	FG01924.1	A0A098D5A0	F9FPC4	2.3.1.7	Carnitina acetiltransferase
16	FG00218.1	IIR9R3	F9G7W3	3.4.13.19	Dipeptidase de membrana NA ⁴
17	FG01017.1	IIRBS7	F9F704	3.4.14.4	Dipeptidil-peptidase III NA
18	FG02492.1	IIRFL3	F9FX31	2.1.1.286	25S rRNA (adenina 2142-N1) -metiltransferase NA
19	FG03349.1	IIRHT5	F9FQP5	1.13.11.1	catecol-1,2- dioxigenase NA
20	FG03624.1	A0A1I9FAY2	F9FI56	3.2.1.8	Endo-1,4-beta- xilanase NA
21	FG03990.1	IIRJH0	F9FYN8	3.1.3.84	ADP-ribose 1 " - fosfato fosfatase NA

22	FG06477.1	I1RQX6	F9G5J4	4.2.1.94	Scitalone desidratase	NA
23	FG07016.1	I1RSA2	F9FMM7	3.2.1.101	Endo-1,6-alfa- manosidase	NA
24	FG10387.1	I1S0Z5	F9FSX4	3.2.1.101	Endo-1,6-alfa- manosidase	NA
25	FG07974.1	I1RUS1	F9FVE8	3.2.1.40	Alfa-L-ramnosidase	NA
26	FG08289.1	A0A0E0RZB4	F9FEN7	3.4.24.39	Deuterolisina	NA
27	FG09073.1	I1RXK2	F9FZ86	3.2.1.40	Alfa-L-ramnosidase	NA
28	FG09485.1	I1RYM8	F9FL57	5.5.1.5	Carboxi-cis, cis- muconato ciclase 25S rRNA (uracil	NA
29	FG09999.1	I1RZY9	F9FI91	2.1.1.312	2843-N3)- metiltransferase	NA
30	FG10343.1	I1S0V3	NA	3.4.14.9	Tripeptidil-peptidase I	NA
31	FG10999.1	A0A1I9FXT9	F9FP27	3.2.1.8	Endo-1,4-beta- xilanase	NA
32	FG11037.1	I1S2N7	NA	3.2.1.151	Endo-beta-1,4- glucanase específica para xiloglucano	NA
33	FG11140.1 ³	A0A098DYT2	F9FPS4	4.1.1.6	Aconitato descarboxilase	NA

Das 30 enzimas atribuídas a 25 EC de *F. oxysporum*, 15 enzimas representadas por 12 ECs (1.13.11.1, 2.1.1.286, 2.1.1.312, 3.1.3.84, 3.2.1.101, 3.2.1.40, 3.2.1.8, 3.4.13.19, 3.4.14.4, 4.1.1.6, 4.2.1.94 e 5.5.1.5) não foram associados a quaisquer vias metabólicas do KEGG.

4.3 Validação das Enzimas Específicas por Análise de Balanço de Fluxo

– FBA

4.3.1 Análise de balanço de fluxo da reconstrução metabólica de *F.*

oxysporum

A reconstrução metabólica de *F. oxysporum* realizada por Pitkänen et al. (293) composta de 882 genes, cujas funções destes genes foram transferidas por homologia de sequência entre 49 fungos com o banco de dados do SwissProt através de BLAST e da ferramenta *Global Trace Graph* (GTG) (299) foi atualizada com 6 novos genes representando 6 novas atividades enzimáticas provenientes do objetivo que visou identificar enzimas específicas entre *F. oxysporum* e cinco espécies de plantas, incluindo *A. thaliana* (Tabela 12). Ao final, a reconstrução metabólica de *F. oxysporum* possuiu 888 genes.

Tabela 12: Novos genes e atividades enzimas inseridas na Reconstrução metabólica de *F. oxysporum*.

#	FOX acesso (UniProtKB)	EC	Nome
1	F9FYW8	2.7.7.53	ATP adenililtransferase
2	F9F7R0	4.3.2.3	Ureidoglicolato liase
3	F9FE93	3.6.1.59	m7GpppX difosfatase
4	F9FX31	2.1.1.286	25S rRNA (adenina 2142-N1) -metiltransferase
5	F9FL57	5.5.1.5	Carboxi-cis, cis-muconato ciclase
6	F9FPS4	4.1.1.6	Aconitato descarboxilase

PORCENTAGEM DOS COMPONENTES DE BIOMASSA

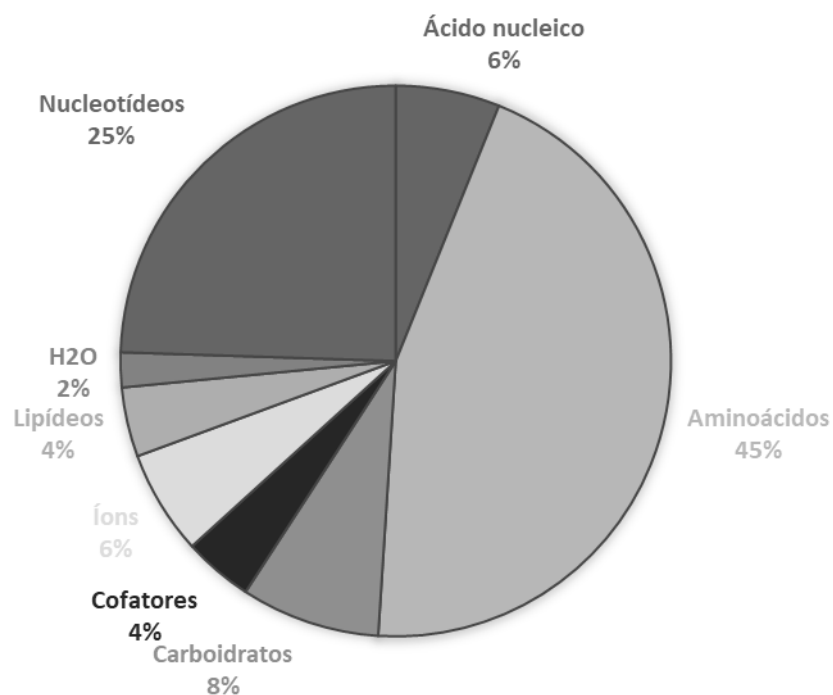


Figura 7: Porcentagem dos componentes de biomassa.

A contribuição de cada gene para a produção de biomassa foi avaliada pela sua exclusão da matriz RPG. Para este fim, cada gene de *F. oxysporum* foi mapeado para sua atividade enzimática usando a anotação KEGG.

Um total de 966 atividades enzimáticas foram mapeadas em 888 genes de *F. oxysporum*. Nenhum dos 6 genes específicos identificados como sendo específicos de *F. oxysporum* em relação as cinco espécies de plantas mencionadas no objetivo anterior foi considerado crítico de acordo com a produção de biomassa sob FBA. No entanto, 39 (4,32%) genes foram considerados críticos em relação à produção de biomassa por *F. oxysporum*.

4.3.2 Detecção de homologia entre genes de *F. oxysporum* e *A. thaliana*

A partir da lista de genes críticos gerados na seção anterior, definimos como alvos potenciais aqueles que não possuem homólogos próximos no genoma do hospedeiro. A pesquisa de similaridade foi realizada entre as sequências de aminoácidos UniProt para (a) genes críticos da reconstrução de *F. oxysporum* e (b) todos os genes incluídos no AraGEM, o modelo metabólico mais curado de *A. thaliana* (300). Os pares homólogos foram definidos como melhores resultados recíprocos BLASTp com $E\text{-value} \leq 10^{-4}$ e uma pontuação de $score \geq 120$.

Foram encontrados três genes que podem servir como alvos para o desenho de fármacos: B0SYU4, F9F4G5 e F9FSB6. O primeiro foi incluído na reconstrução de *F. oxysporum* para *gap-filling* e não tem homólogos próximos no genoma do organismo e pode não ser uma escolha confiável. Os outros dois, F9F4G5 (EC: 3.5.4.26) e F9FSB6 (EC: 3.5.1.49), têm análises funcionais em *A. thaliana*, Q8GWP5 (EC: 3.5.4.26) e Q9SZE9 (EC: 3.5.1.49), respectivamente e devem ser analisados para além do nível da estrutura primária, uma vez que não há tradução direta entre a sequência de aminoácidos e a estrutura 3D, que é a informação final necessária para finalmente decidir se os pares de enzimas podem ser considerados análogos ou não.

4.3.3 Modelagem tridimensional das duas enzimas análogas consideradas críticas

Como mostrado acima, a enzima pseudouridina sintase, F9F4G5, de *F. oxysporum* compartilha uma relação análoga com a enzima proteína de biossíntese de riboflavina, Q8GWP5, de *A. thaliana* e, portanto, é um candidata interessante para ser um alvo

adequado para um inibidor específico. No entanto, a hipótese de analogia entre F9F4G5 e Q8GWP5 deve ser confirmada por um método independente para que F9F4G5 possa ser considerado como um alvo relevante. O método independente preferido neste caso é a modelagem *in silico* 3D.

F9F4G5 (monómero) tem um total de 514 aminoácidos e mostrou uma identidade de 25% com o modelo *Escherichia coli* 1V9F (PDB ID) de resolução 1.99 Å, conforme medido por difração de raios X. O modelo gerado tem cobertura de tamanho de 70% e se estende entre os aminoácidos 109 até 466 (Figura 7A). Somente a região central poderia ser modelada; as outras regiões terminais não apresentaram nenhuma informação que permitisse a modelagem por outras técnicas (*ab initio* ou *fold-cognition*) além daquela já utilizada. Q8GWP5 (homo-dímero) possui um modelo 3D de *Acinetobacter baumannii* (PDB ID: 3ZPG) com uma resolução de 1,7 Å medida por difração de raios X e superior aos aminoácidos 73 e 425 de um total de 426 aminoácidos (83% de cobertura) (Figura 8A).

O modelo de *F. oxysporum* possui um bom resultado de validação com uma porcentagem de rotâmeros medidos superior a 90% (Figura 7B). Este resultado é importante dado o baixo nível de identidade (25%) para a modelagem de homologia. No entanto, os outros escores padrão (Procheck 3 (301), Verify3D (302), 5 Prosa II (303), MolProbity Clashscore (304)) (Tabela 13) também são favoráveis ao uso do modelo para outras análises. No modelo de *A. thaliana*, com uma identidade ligeiramente superior (35%), a porcentagem de ângulos favoráveis é de 91% e, assim como a sequência de *F. oxysporum* apresenta valores de validação favoráveis (Tabela 14).

Tabela 13: Escores de análise do modelo de *F. oxysporum*.

	Verify3D	ProsaII(-ve)	Procheck	Procheck	MolProbity
			(phi-psi)	(all)	Clashscore
-Raw score	0,27	0,09	-0,40	-0,33	188,58
Z-score	-3,05	-2,32	-1,26	-1,95	-30,83

Tabela 14: Escores de análise do modelo de *A. thaliana*.

	Verify3D	ProsaII(-ve)	Procheck	Procheck	MolProbity
			(phi-psi)	(all)	Clashscore
-Raw score	0,36	0,35	-0,23	-0,17	166,00
Z-score	-1,61	-1,24	-0,59	-1,01	-26,96

O potencial eletrostático de ambos os modelos gerados apresenta uma enorme diferença, sendo antagonista um ao outro (Figura 7 e 8, B – Direita); um é principalmente positivo, enquanto o outro é principalmente negativo. Esses recursos se assemelham aos seus respectivos moldes, dando ainda maior relevância aos dados obtidos. As características hidrofóbicas de ambos não apresentam o mesmo alto grau de diferença, no entanto, ainda são diferentes.

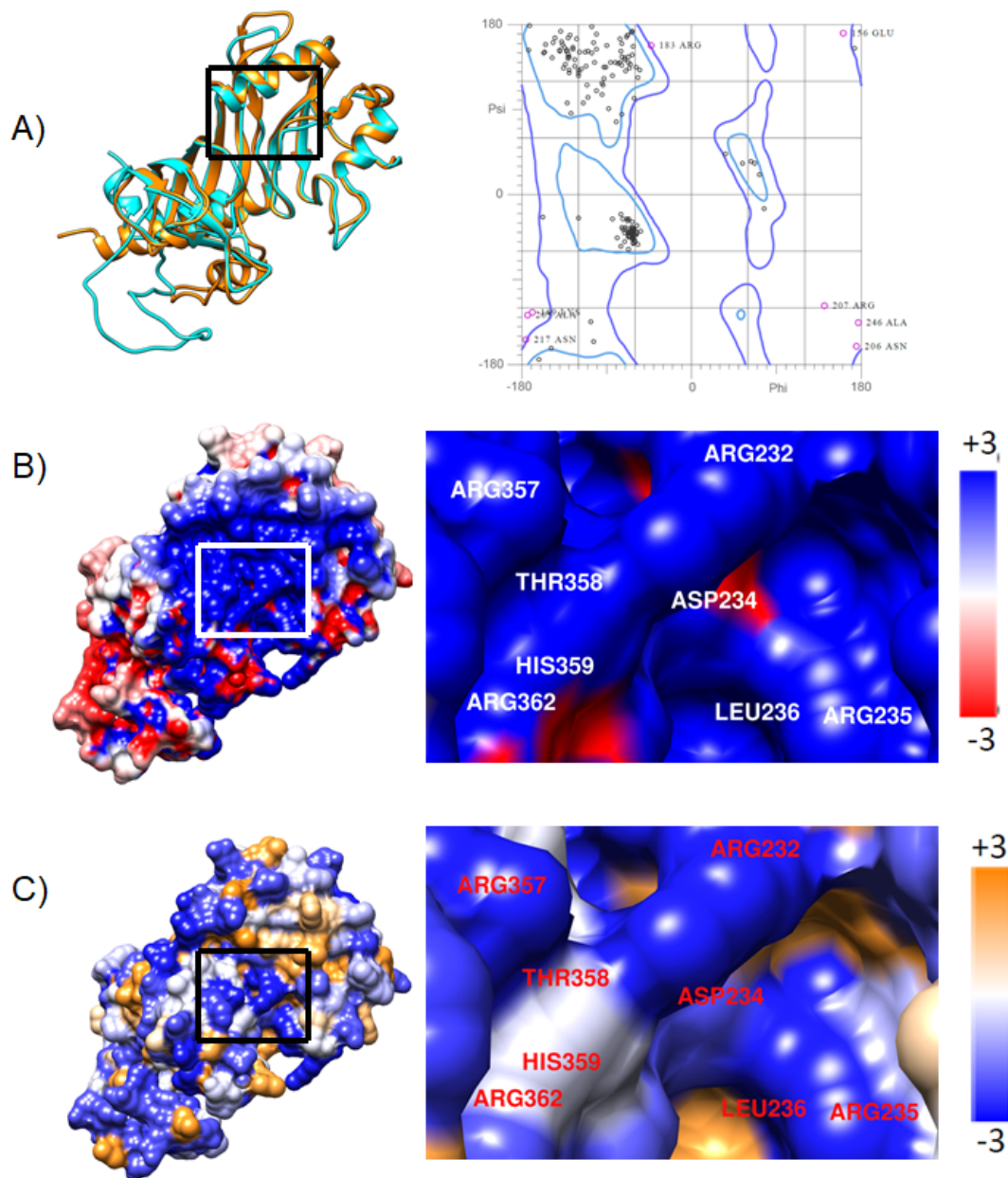


Figura 8: Modelagem 3D da enzima Pseudouridina sintase (EC:3.5.4.26). **A)** esquerda: Estrutura monomérica da enzima Pseudouridina sintase (EC: 3.5.4.26) de *F. oxysporum* (ciano), sobreposta à estrutura de *E. coli* (APB: 1V9F - laranja). Direito: Gráfico de Ramachandran com porcentagem favorável de 90%. **B)** Esquerda: potencial eletrostático (Kb.T.ec^{-1}) - azul: negativo, vermelho: positivo, toda a estrutura. Direito: potencial eletrostático do sítio ativo. **C)** Hidrofobicidade - laranja: hidrofóbico; azul: hidrofílico. Esquerda: toda a estrutura. Direita: foco do sítio ativo.

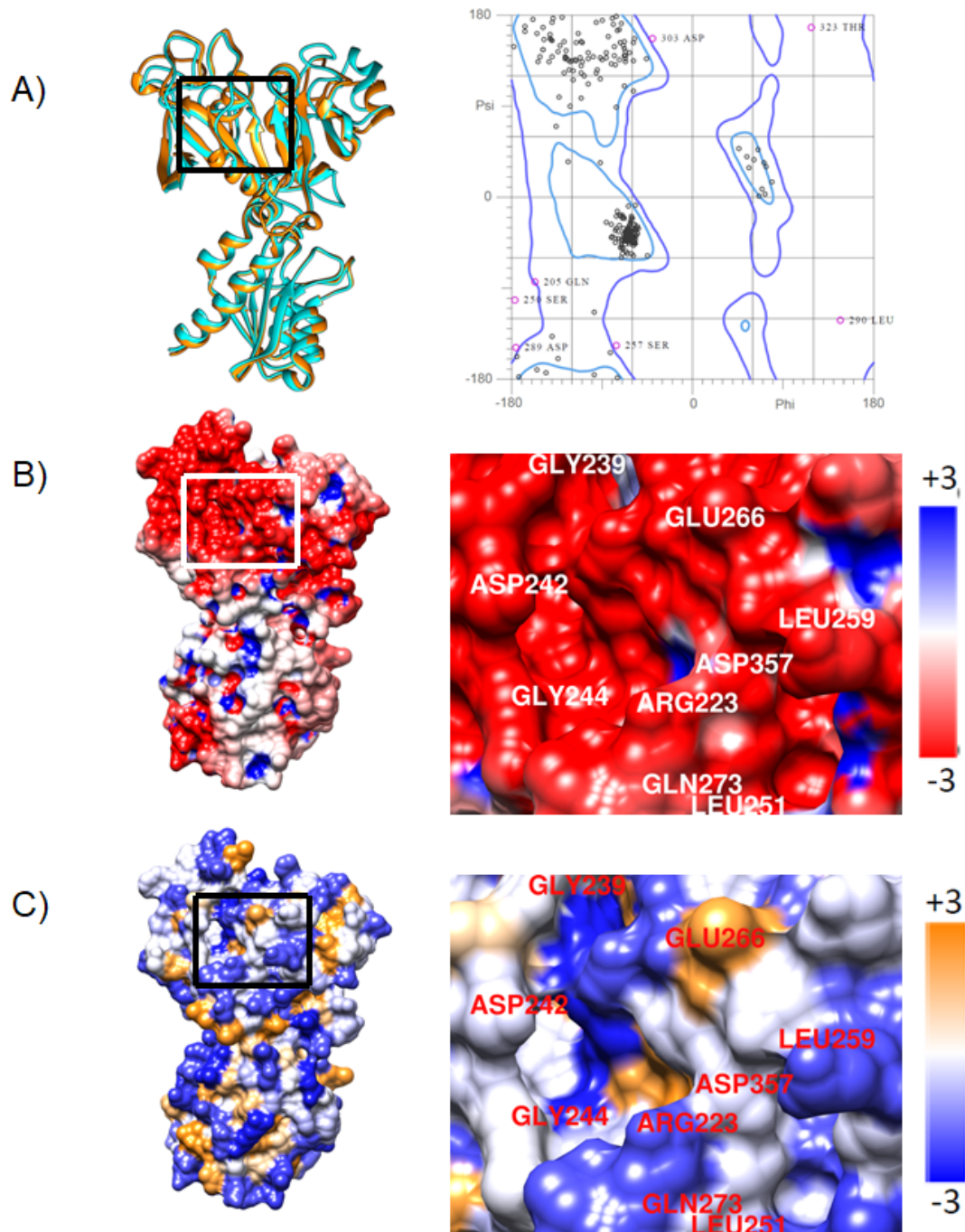


Figura 9: Modelagem 3D da enzima responsável pela biossíntese de Riboflavina (EC: 3.5.4.26). **A)** Esquerda: Estrutura monomérica da biossíntese de Riboflavina (EC:3.5.4.26) de *A. thaliana* (ciano), sobreposta à estrutura de *A. baumannii* (PDB: 3ZPG - laranja). Direito: Gráfico de Ramachandran com porcentagem favorável de 91%. **B)** Esquerda: Potencial

eletrostático ($Kb.T.ec^{-1}$) - azul: negativo, vermelho: positivo, toda a estrutura. Direito: potencial eletrostático do sítio ativo. C) Hidrofobicidade - laranja: hidrofóbico; azul: hidrofílico. Esquerda: toda a estrutura. Direita: foco do sítio ativo.

5. DISCUSSÃO

5.1 Busca de Enzimas Específicas em *L. major*

5.1.1 Desafios de classificação e ambiguidade de anotação

A primeira parte da tese teve como objetivo descrever um processo para filtrar as anotações falso-positivas de especificidade enzimática nas interações hospedeiro-parasito produzidas pelo AnEnPi (37). A identificação de alvos proteicos específicos em parasitos (*L. major*) em comparação com o hospedeiro (*H. sapiens*) é importante para o desenvolvimento de medicamentos com o menor impacto por efeitos colaterais possíveis sobre o hospedeiro. O procedimento apresentado é sistemático, compatível com a automação e adequado para interações hospedeiro-parasitos que envolvem eucariotas inferiores como parasitos e eucariotas superiores como hospedeiros. O fato de que os alvos identificados por esta metodologia já foram previamente identificados sugere que os resultados são válidos e aumenta a nossa confiança de que a metodologia proposta pode ser aplicada a outros sistemas hospedeiro-parasito com menor nível de conhecimento associado.

Foram analisadas 67 sequências de proteínas que correspondem a 42 atividades enzimáticas e essas sequências foram previamente classificadas como específicas de *L. major* em comparação com *H. sapiens* pelo AnEnPi. Em uma comparação criteriosa utilizando homologias dessas sequências com as sequências do Ensembl utilizando o BLAST permitiu identificar 15 pares homólogos significativos entre *L. major* e *H. sapiens*. Estes pares homólogos foram associados a ECs diferentes em *L. major* e *H. sapiens*, sugerindo que estes deveriam ser classificados como funcionalmente específicos em *L. major* em comparação com *H. sapiens*. É claro que a classificação das

enzimas como sendo específicas funcionais é completamente dependente das anotações de EC, e isso pode resultar em erros triviais, como é o caso das ATPases que podem transportar cátions H^+ ou Na^+ através das membranas celulares. As 2 ATPases carregadoras de H^+ ou Na^+ são homólogas e a diferença em seus números EC está associada aos cátions transportados (a reação e não a atividade enzimática), que não estão associados à própria atividade ATPase quando o substrato é trifosfato de adenosina (ATP). Assim, deve-se ter cuidado ao interpretar ECs. Nos outros casos, apenas encontramos diferenças mínimas ao considerar os perfis das estruturas 2D entre os membros da sequência dos pares homólogos provenientes do BLAST.

Entre as 15 enzimas funcionalmente específicas putativas, os ECs de 9 enzimas de *L. major* diferiram apenas das enzimas de *H. sapiens* no quarto dígito. Ao comparar os alinhamentos 2D desses 9 pares homólogos, apenas encontramos pequenas diferenças, e é difícil afirmar que um fármaco para os alvos em *L. major* não teria algum tipo de interação indesejada com a forma humana o que poderia resultar em efeitos colaterais negativos para os pacientes.

O banco de dados Ensembl envolve processos de anotação automática de genes que eventualmente passam por processos de curadoria manual; este é também o caso de Havana (305,306). O *pipeline* do AnEnPi é completamente automático, e classifica as enzimas como específicas, homólogas ou análogas de acordo com suas atividades enzimáticas e homologias. Obviamente, anotações enzimáticas incompletas levam a classificações não otimizadas pelo AnEnPi. O objetivo deste estudo foi detectar esses casos e fornecer mais conhecimento para melhorar o desempenho geral do classificador automático AnEnPi.

Embora os métodos de sequenciamento tenham produzido grandes quantidades de dados, as montagens de genomas com base nos dados produzidos podem estar incompletas ou aproximadas (307). As montagens (*assemblies*) de baixa qualidade resultam em anotações de baixa qualidade (308,309) e promovem subestimações sobre o número de genes (310). Considerando o organismo em questão em uma comparação, a montagem incompleta do seu genoma pode levar à perda de genes e assim gerar genes específicos virtuais no genoma da consulta. É possível verificar a consistência de um evento que é observado, mas é quase impossível verificar a consistência de um evento que não é observado. Assim, neste estudo, não foram considerados casos de genes perdidos por causa de erros de montagem como fonte de falsos positivos. Apenas uma análise *in vitro* por inativação do conjunto de enzimas putativamente específicas proposto pode resolver esta questão.

Por exemplo, a existência de pseudogenes pode ser virtual ou real, dependendo se eles são a consequência de um erro no processo de sequenciamento (mudanças de quadros induzidas e codons de parada prematura) ou processos de montagem (corte de genes) que levam a anotações imprecisas (311–313). De fato, muitos pseudogenes verdadeiros têm alguma forma de atividade biológica e, portanto, sua anotação precisa é potencialmente importante para o desenvolvimento de fármacos (313). Observamos apenas 5 casos de pseudogenes humano associados às proteínas de *L. major*.

As homologias obtidas pela implementação de HHpred e HHsearch têm um nível de confiança maior em comparação com o BLAST, pois os HMMs consideram a similaridade nos pares de sequências primária de aminoácidos a partir da sequência *query* de *L. major* e da sequência humana bem como as estruturas 2D e 3D de cada sequência.

5.1.2 Essencialidade enzimática

A atividade enzimática pode desempenhar um papel fundamental em uma via metabólica. Cada rota está associada a enzimas e seus ECs. Portanto, uma via metabólica é um gráfico orientado cujos vértices são caracterizados por uma enzima e sua reação química associada (314). Uma enzima chave é uma enzima que está contida em um caminho, não possui uma opção alternativa e desempenha um papel essencial na sobrevivência de um organismo. Consequentemente, a inibição de uma enzima chave necessariamente resulta na inibição da via correspondente e da debilitação do organismo. No entanto, se um alvo está em uma rota que não desempenha um papel importante na sobrevivência do organismo, a inativação deste alvo não acarreta a inibição das funções vitais do organismo e ele pode continuar a se reproduzir. Assim, a característica chave para explorar uma via metabólica é identificar as enzimas que desempenham papéis essenciais para a sobrevivência do parasito, ou seja, cuja inibição é prejudicial para esse organismo. No contexto do desenvolvimento de medicamentos, o medicamento ideal deve ser o que pode inibir a atividade de uma enzima específica em uma via essencial do organismo alvo, sem rota alternativa.

5.1.3 Especificidade funcional

A especificidade estrita, ou seja, a existência de uma determinada enzima no parasito e não no hospedeiro, é, obviamente, o indicado. No entanto, a especificidade funcional, em que a especificidade do substrato pode ser identificada no parasito em comparação com a forma do hospedeiro, também pode ser contemplada. Na verdade, o

EC: 1.3.1.71 (*L. major*), que está associado à via de biossíntese de esteroides, é um exemplo de enzima funcionalmente específica que é homóloga à enzima humana que possui o EC: 1.3.1.70. O EC: 1.3.1.71 poderia ser explorado para o desenvolvimento de fármacos, como agentes antifúngicos e anti-tripanosomatídeos (315). Xu et al. (316) mostraram que a eliminação do gene que codifica a enzima 14- α -desmetilase (EC:1.14.13.70), que é anterior ao EC: 1.3.1.71, em *L. major*, contribuiu inicialmente a uma mudança dramática no perfil da composição lipídica em amastigotas e à atenuação da virulência, mas estes efeitos foram revertidos após algumas semanas. Como a atividade de EC:1.14.13.70 é comum à via de sínteses de colesterol e ergosterol em mamíferos e *Leishmania* spp, respectivamente, é possível que *L. major* tenha contornado a perda da função do EC:1.14.13.70 nos parasitos após inativação (*knockdown*) usando esteróis do hospedeiro. Esta estratégia de escape seria pouco provável se o EC:1.3.1.71 fosse inibido em vez da atividade da enzima 14- α -desmetilase em *Leishmania* spp, já que EC:1.3.1.71 é inserido em uma via de síntese de ergosterol que é específica para Trityps. Não foi possível encontrar outros casos funcionalmente específicos, potencialmente confiáveis, neste estudo, pois podem existir rotas de metabólitos alternativas para o parasito para sustentar a função que foi inibida pelo tratamento medicamentoso (316). Mais recentemente, McCall et al. (317) demonstraram a importância da enzima 14- α -desmetilase em *L. donovani*.

5.1.4 Especificidade estrita

As enzimas estritamente específicas podem ser divididas em alvos com um menor (32 enzimas caracterizadas como sendo “secundárias”) e um maior (8 enzimas primárias) potencial para o desenvolvimento de fármacos.

Considerando as enzimas estritamente específicas com maior potencial para desenvolvimento de fármacos, a análise realizada mostrou que 4 funções enzimáticas, esterol 24-C-metiltransferase (EC:2.1.1.41), tripanotiona sintetase (EC:6.3.1.9), piruvato fosfato diquinase (EC:2.7.9.1) e a RNA ligase (EC:6.5.1.3), podem ser consideradas para o desenvolvimento de um coquetel de fármacos.

Ao contrário dos humanos que possuem colesterol em suas membranas biológicas, os tripanosomátideos usam ergosterol. Portanto, os tripanosomátideos apresentam enzimas distintas na sua via de biossíntese de esterol (318). A enzima esterol 24-C-metiltransferase (EC:2.1.1.41) catalisa uma metilação do carbono 24, que é fundamental para a biossíntese de ergosterol. A reação 24-C-metiltransferase é inibida pelo 22,26-azasterol, que causa mudanças morfológicas e lise em *Leishmania* spp (315).

A tripanotiona sintetase (EC 6.3.1.9) é fundamental para a síntese de tripanotiona, um metabólito de tripanosomátideos que está envolvido em vários processos, como a regulação do equilíbrio redox de tiol intracelular, resistência a fármacos, defesa contra agentes químicos e estresse oxidativo (319). As enzimas tripanotiona sintetase e tripanotiona redutase estão envolvidas nas vias de biossíntese e metabolismo da tripanotiona e estão em estudo para o desenvolvimento de uma quimioterapia alternativa contra tripanosomátideos (320).

A enzima piruvato fosfato diquinase (EC:2.7.9.1) é uma enzima presente nos glicossomos dos tripanosomátideos (321). Foi relatado que esta enzima desempenha um papel fundamental na manutenção do equilíbrio de ATP vs difosfato de adenosina nas organelas da forma pró-cíclica de *T. brucei* (322) e no processo de gliconeogênese nas formas amastigotas de *L. mexicana* (323). Portanto, acredita-se que a via que inclui a enzima piruvato fosfato diquinase é central para o metabolismo energético de *L.*

major e que esta enzima merece ser considerada como um alvo para o desenvolvimento de fármacos leishmanicidas.

O complexo L foi isolado das mitocôndrias de *T. brucei* e *Leishmania tarentolae*. Pelo menos 16 componentes do complexo foram identificados, incluindo REL1 e REL2. A função precisa de REL1 e REL2, ambas anotadas como EC:6.5.1.3, foi descrita por complementação de cepas *knockout* (324,325). Em *T. brucei*, a interrupção condicional de REL1 é letal *in vivo*, pois afeta a edição tanto da deleção de uracila quanto a inserção resultando em uma diminuição geral no tamanho do RNA (326). No entanto, a perda de REL2 não tem efeito na viabilidade ou na edição (327).

Estas enzimas atuam em diferentes rotas metabólicas e processos celulares: biossíntese de esterois e estrutura da membrana (EC:2.1.1.41), metabolismo da glutatona e resposta ao estresse oxidativo (EC:6.3.1.9), metabolismo energético e via de fixação de carbono no metabolismo de procariotos (EC:2.7.9.1) e edição de RNA / processamento pós-transcricional de RNA (EC:6.5.1.3). Portanto, a inibição simultânea dessas enzimas e conseqüentemente essas vias / processos celulares provavelmente poderão promover um colapso irreversível das células de *L. major*.

Usando uma estratégia baseada em drogabilidade de alvos, Crowther et al. (328) identificaram as enzimas LmjF.27.1870 (EC:6.3.1.9), LmjF.36.2380 e LmjF.36.2390 (as duas últimas tendo EC:2.1.1.41) como possíveis alvos para desenvolvimento de fármacos o que vai de acordo com resultados encontrados neste trabalho. No entanto, essas enzimas foram posteriormente desconsideradas, pois não foram classificadas como ortólogas às enzimas classificadas como essenciais em outros organismos modelo (*Caenorhabditis elegans*, *Escherichia coli*, *M. tuberculosis* e / ou *Saccharomyces cerevisiae*) através de experimentos de bancada, sendo que talvez essas enzimas não

foram classificadas como essenciais, pois não existiam nesses organismos mencionados, conforme demonstrado na Tabela 6 para *M. tuberculosis*.

Em contraste, com base na inferência do interactoma, Flórez et al. (329) identificaram 142 alvos específicos na rede de sinalização de *L. major* em comparação com *H. sapiens*. O único EC entre os nossos 8 alvos estritamente específicos que foi possível encontrar em comum com sua lista foi EC:2.1.1.41 (LmjF.36.2380 e LmjF.36.2390), o que parece contrariar a afirmação de Crowther et al. (328). Além disso, a enzima esterol 24-C-metiltransferase foi identificada como um alvo promissor para desenvolvimento de fármacos em *Cryptococcus neoformans* (330). Goto et al. (331,332), também, descreveram uma vacina candidata para a enzima esterol 24-C-metiltransferase contra a leishmaniose visceral que é eficaz contra *L. infantum* e *L. donovani*.

Ao contrário de Crowther et al. (328) a essencialidade da enzima foi avaliada, através de sua centralidade em uma via metabólica no caso específico de *L. major* usando sua representação metabólica no KEGG.

Assim, as 8 enzimas estritamente específicas representadas pelos 4 ECs que consideramos essenciais são também relatadas como essenciais na literatura. Contudo, investigando o TDR (<http://tdrtargets.org/>), foi possível notar que 5 enzimas são consideradas essenciais para um ou mais tripanossomatídeos e 1 (LmjF.36.4300) não possui dados sobre sua associação de essencialidade (333).

5.1.5 Redundância genética e resistência parasitária aos fármacos

Foi possível verificar que, com exceção do EC:2.7.9.1, que possui apenas 1 sequência enzimática associada, os outros 3 números de EC para alvos estritamente específicos (EC:2.1.1.41, EC:6.5.1.3 e EC:6.3.1.9) têm mais de 1 gene representativo no genoma de *L. major*. A redundância genética para uma mesma função enzimática (ver Tabela 8) é uma fonte potencial de resistência a fármacos, porque um dado inibidor pode inibir uma determinada enzima, mas não suas isoformas putativas, que também podem estar presentes (334). A resistência a medicamentos pode envolver vários mecanismos diferentes, e uma solução comum para este problema é aumentar o número de enzimas que são alvo pelos inibidores ao mesmo tempo.

O processo de acumulação de genes, que foi chamado de piramidação por Bury et al. (335), é rotineiramente usado no melhoramento genético clássico vegetal para induzir resistência a agentes patogênicos, onde foi implementado pela primeira vez (336,337). A piramidação reduz a probabilidade de adaptação por um parasito, pois o acúmulo correspondente de genes de compatibilidade torna-se insustentável nas condições ambientais fornecidas.

A ideia de acumular uma série de alvos que são simultaneamente inibidos para um controle melhor de doenças é uma variação da relação gene-para-gene que foi descrita por Flor (338) e pode ser referida como o conceito de gene para inibidor (339). Assim, a formulação de fármacos em um coquetel deve superar a resistência parasitária; esta prática (340) já se encontra em uso. No entanto, a formulação de uma combinação de fármacos também deve minimizar os efeitos colaterais negativos às células normais e proteger a integridade do sistema imune do hospedeiro limitando a dose de fármacos administrados. Assim, em uma tentativa de abordagem desta questão foram

identificados apenas os alvos enzimáticos que são específicos para *L. major*, já que visando esses candidatos provavelmente minimizaria os efeitos colaterais deletérios de uma combinação terapêutica aos pacientes. Claro, que os resultados delineados neste trabalho ainda devem ser confirmados por experimentação *in vitro*. Outra questão a ser avaliada no futuro é a exigência de estudos de segurança para cada um dos componentes do tratamento combinado, independentemente e em combinação, antes do registro bem-sucedido desse tratamento combinado; no entanto, esta questão está além do escopo deste trabalho, que é principalmente exploratório.

5.1.6 Fármacos para enzimas de *L. major* estritamente específicas

O reposicionamento de medicamentos é uma rota preferencial no desenvolvimento de fármacos devido à sua economia de tempo e dinheiro. Infelizmente, ao examinar o *TDR Targets* não foi possível obter resultado de nenhum fármaco aprovado para os alvos enzimáticos que foram descritos, mas alguns inibidores parecem estar disponíveis pelo menos para os estádios experimentais em *Leishmania* ou outras espécies biológicas. Pesquisando a literatura descobriu-se que (i) os azasteróis são inibidores da esterol 24-C-metiltransferase (EC:2.1.1.41) em *Leishmania* (315,340); e (ii) corante de azoícos do tipo naftaleno, semelhantes à suramina, são inibidores de REL (EC:6.5.1.3) (341). Dois desses compostos foram posteriormente testados em todo o editosomo e mostraram forte inibição da deleção de U na edição de RNA (342). GW5074, mitoxantrona, NF 023, protoporfirina IX e d-esfingosina também foram identificados como inibidores de REL, com valores de IC₅₀ que variam de 1 a 3 µM (343); (iii) o fosfoenolpiruvato (ID: DB02522 no *DrugBank*) foi caracterizado como um inibidor (344) de piruvato fosfato diquinase (EC:2.7.9.1), bem como análogos de

flavonas. Também, foi relatado que os análogos de flavona atuam especificamente visando o sítio de ligação ao ATP da enzima em um sistema de fungos (157). Vários compostos fenólicos também foram propostos para a inibição da piruvato fosfato diquinase em *Wolbachia* (345); (iv) Um derivado do bis-benzil-diamina, bem como MOL2008, foi proposto como novos *scaffolds* moleculares para o desenvolvimento de fármacos, que provaram ser ativos em $EC_{50} = 12 \mu\text{M}$ em promastigotas de *L. infantum*; estes compostos eliminaram a tripanotona intracelular através da inibição de tripanotona sintetase (EC:6.3.1.9) em parasitos tratados, confirmando a atividade alvo desses compostos (346).

Sob a hipótese de que a susceptibilidade de *Leishmania* seria estável *in vitro* para uma combinação dos compostos listados acima, seria possível otimizar sua estrutura de *scaffold* para reduzir a toxicidade geral do coquetel para um modelo de célula humana. A engenharia deste processo agora é facilitada pela modelagem *in silico*. Uma pesquisa de modelos 3D no PDB com sequências homólogas a LmjF.36.2380, LmjF.36.2390, LmjF.11.1000, LmjF.23.0460, LmjF.27.1870, LmjF.36.4300, LmjF.01.0590 e LmjF.20.1730 mostrou que apenas as sequências possuidoras dos EC:2.7.9.1 e EC:6.3.1.9 eram adequadas para a modelagem 3D, porque os seus índices de identidade eram maiores que 39%, chegando a um nível máximo de 97%. No entanto, no caso dos EC:6.5.1.3 e EC:2.1.1.41, os índices de identidade eram muito baixos (<25%) e o número de *gaps* era muito alto para modelagem 3D bem sucedida.

5.1.7 Distribuição de enzimas estritamente específicas de *L. major* entre parasitos humanos

A enzima esterol 24-C-metiltransferase (EC:2.1.1.41) é um componente da via do ergosterol. Esta via tem sido amplamente estudada para o desenvolvimento de medicamentos contra fungos, e a inativação da via do ergosterol demonstrou-se como eficiente e estável. O fato de que a esterol 24-C-metiltransferase poder ser compartilhada com fungos significa que a reposição de fármacos de fungos para *Trypanosomatidae* pode ocorrer. O piruvato fosfato diquinase (EC:2.7.9.1) é altamente específico para *Trypanosomatidae*, e o EC:2.7.9.1 só foi encontrado na micobactéria *M. tuberculosis*; no entanto, as sequências não se alinharam, o que sugere que as enzimas de ambas as linhagens são análogas. As sequências de tripanotiona sintetase (EC:6.3.1.9) de *T. cruzi* apresentaram identidades entre 37% e 58% com as de *L. major*. Esses dados estão de acordo com a literatura, o que demonstra a expressão desta enzima e a presença desta no metabolismo de tripanotiona em *T. cruzi* (328). Como esperado, *L. major*, que pertence ao Tritryps, apresentou maior afinidade filogenética com *T. cruzi* e *T. brucei* (Euglenozoa) do que com *P. falciparum*, que pertence a um filo diferente (Apicomplexa). O status da analogia de sequência que foi atribuído ao EC:6.5.1.3 (REL1) de *P. falciparum* e *T. gondii* não é surpreendente, considerando que ambas as espécies pertencem ao clado de Alveolatos (Chromalveolates), enquanto *Trypanosomatidae* pertence a Euglenozoa (Excavates), que estão tão distantes um do outro, em termos evolutivos, como são distantes do filo Animalia (Unikonts) (<http://tolweb.org/Eukaryotes/3>). No entanto, o estado da analogia depende apenas da anotação do EC. A esse respeito, pode-se perguntar se uma ligase de edição de RNA que está envolvida no processo de *trans-splicing* em Tritryps deve ter o mesmo EC (EC:6.5.1.3) que uma ligase de tRNA-*splicing* que está envolvida no reparo de cis-

splicing e tRNA, uma vez que as consequências biológicas das 2 reações são diferentes em ambos os casos. Curiosamente, REL1 é conservada em toda a subclasse de Metakinetoplastina, que mostra o seu antigo surgimento dentro de Excavates. Infelizmente, não encontramos uma sequência genômica completa para nenhum membro do filo Diplonema ou Euglena de posição taxonômica superior para verificar o nível preciso da história da vida em que a enzima REL surgiu. Além dos parasitos, os Metakinetoplastina também incluem formas de vida livre não-parasitos, como *B. saltans*, que mostra que a função REL é anterior à adaptação parasitária. Na verdade, a REL é codificada por um gene informacional e pertence às funções primordiais que foram usadas pela vida assim que o mundo do RNA surgiu, (347) e foi demonstrado que foi inventada pelo menos 2 vezes de forma independente.

5.1.8 Padrão de expressão das principais enzimas específicas de *L. major*

A expressão de LmjF.01.0590, LmjF.11.1000, LmjF.20.1730, LmjF.23.0460, LmjF.27.1870, LmjF.36.2380, LmjF.36.2390 e LmjF.36.4300 é descrita como sendo constitutiva na forma promastigota (sangue) e amastigota (intracelular) de *L. major*, (348–350) o que torna estas sequências alvos adequados para fármacos que sejam eficazes contra ambas as formas de parasitos.

5.2 Busca de Enzimas Específicas em *F. oxysporum*

A toxicidade causada ao meio ambiente por fungicidas é certamente um motivo para busca de alvos enzimáticos específicos em pragas em geral e em *F. oxysporum* em particular. Culturas como *B. rapa*, *G. max*, *J. curcas* e *R. communis* deverão ter um

grande impacto no meio ambiente devido à sua importância comercial no contexto da produção de biodiesel. A planta *A. thaliana* é um modelo para essas espécies de plantas, o que justificou a escolha desta no trabalho.

A metodologia desenvolvida na seção 3.1 relativa ao modelo *H. sapiens* vs. *L. major* está aqui aplicada em outro sistema parasito-hospedeiro, neste caso, fungo-planta e consistiu tanto na possibilidade de validação da aplicabilidade da metodologia em outros sistemas parasito-hospedeiro quanto na possibilidade de buscar por alvos enzimáticos específicos de *F. oxysporum* que pudessem ser candidatos para desenvolvimento de novos fungicidas.

Ao comparar a espécie *F. graminearum* com mais de um hospedeiro de plantas, o número de enzimas específicas deste fungo tende a diminuir, o que mostra que cada par de hospedeiro-parasito deve ser considerado em suas próprias características. Na verdade, esse relacionamento não é surpreendente, pois os parasitos coevoluem com seus hospedeiros de forma a manter a virulência e a concorrência com outros micróbios em seu ambiente. Assim, espera-se que os genes mantidos em um par hospedeiro-parasito possam ser diferentes de um par que envolva uma espécie de planta diferente, mesmo que a própria espécies de fungo seja a mesma.

Na verdade, *Fusarium* spp. são fungos ascomicetos haploides de evolução rápida com uma variação de número de genes entre espécies de pelo menos 10% (351). Uma vez que as espécies de plantas envolvidas na interação são diferentes, não é surpreendente encontrar diferença no número de genes para enzimas específicas do fungo, dependendo do sistema hospedeiro-parasito considerado.

Além disso, outros fatores como transferência horizontal de genes (THG) pode estar envolvida na patogênese (352,353) e podem promover troca de informação

genética entre duas espécies (354–357). Richards et al. (199) demonstraram eventos de THG ocorrendo de *Fusarium* para o arroz. E ainda, diversas plantas como *A. thaliana* obtiveram a via do ácido chiquímico (chiquimato) a partir de fungos simbióticos via THG (205,358,359).

De acordo com a indicação de que a inibição de uma enzima pode desativar (ou não) a via metabólica e causar danos irreversíveis (ou não) ao fungo *F. oxysporum* de forma a evitar a sua sobrevivência, as enzimas específicas foram classificadas em enzimas principais (primárias) e enzimas auxiliares (secundárias). As enzimas primárias podem ser consideradas como alvos promissores para o desenvolvimento de novos fungicidas, pois sua inibição tem, em princípio, a capacidade de perturbar a via metabólica nas quais elas estão inseridas e com isso, levar o fungo à morte. Em contraste, as enzimas secundárias possuem um papel auxiliar, pois elas estão inseridas em rotas alternativas e em caso de completa inibição, elas não seriam capazes de causar a morte do fungo.

5.2.1 Enzimas Principais

Podemos enumerar por ordem de prioridade decrescente as enzimas consideradas chave para o controle de *Fusarium* spp. de forma a trazer uma breve descrição de seus papéis fisiológicos que possa justificar sua inibição.

5.2.1.1 Quitina sintase

Em fungos, o principal constituinte da parede celular é o polímero de quitina, que consiste no produto de uma série de reações bioquímicas, onde a primeira consiste na síntese de UDP N-acetilglicosamina (UDP-GlcNAc) pela enzima quitina sintase (EC:2.4.1.16). As células em que a formação de quitina é interrompida são afetadas por sensibilidade osmótica, morfologia anormal, e interrupção do crescimento (360). Embora algumas enzimas da via biosintética de quitina possam servir como alvos no desenho de moléculas antifúngicas, a maioria dos estudos se concentram na enzima quitina sintase (361). Como a quitina é ausente em vertebrados e plantas, a inibição de sua biossíntese pode ser considerada como uma opção segura e seletiva para o desenvolvimento de agentes antifúngicos (362). Logo, a proposta visa desativar a enzima quitina sintase (EC:2.4.1.16) a fim de evitar a síntese de quitina sem danificar a planta. Esta ideia é plausível, pois a quitina sintase é uma enzima específica do fungo presente em uma via metabólica essencial, sem rota alternativa. Uma outra possível forma de inibição do crescimento de *Fusarium* seria através do bloqueio da atividade da enzima UDP-N-acetilglicosamina fosforilase (EC:2.7.7.23), que é responsável pela conversão N-acetilglicosamina-1-fosfato em UDP-GlcNAc, um monômero utilizado pela enzima quitina sintase para a produção de quitina através de polimerização. Entretanto, a enzima UDP-N-acetilglicosamina fosforilase se encontra presente em plantas e de acordo com Jiang et al. (363), Nozaki et al. (364) e Chen et al. (365), a síntese de UDP-GlcNAc possui um impacto direto no crescimento e desenvolvimento de plantas, que torna a inibição do EC:2.7.7.23 uma questão a ser avaliada, uma vez que esta pode causar danos significativos ao hospedeiro.

As nicomicinas e polioxinas foram os primeiros inibidores de quitina sintase descritos (366), porém apesar de excelentes resultados *in vitro*, elas demonstraram uma baixa atividade *in vivo* devido ao seu transporte ineficiente ou degradação antes de alcançar a enzima alvo (367–371). Ge et al., (372) relataram a síntese e avaliação biológica de 20 derivados como possíveis candidatos de inibidores da enzima quitina sintase entre os quais um composto possuiu um resultado promissor ao obter um IC₅₀ de 0.1 mM e, ser então, considerado o composto mais promissor para o futuro desenvolvimento de agentes seletivos antifúngicos, principalmente contra fungos do gênero *Fusarium*.

5.2.1.2 Enzima *decapping scavenger*

A enzima *decapping scavenger* (DcpS, m7GpppX difosfatase, EC:3.6.1.59) é primordial no controle da estabilidade do mRNA e na regulação da expressão gênica (218). A enzima é responsável pela hidrólise da estrutura denominada *cap* que resulta da decomposição da via de degradação 3'-5' (373,374). Alguns inibidores da enzima DcpS foram descritos para o tratamento da atrofia muscular espinhal (AME). Esses compostos inibidores foram derivados da substituição da unidade C5 de quinazolininas e foram capazes de inibir a atividade enzimática de DcpS, sendo que o composto D156844 foi indicado como o inibidor mais potente (375–377).

5.2.1.3 Carnitina O-acetiltransferase

A ligação de unidades de acetil à carnitina (378) ocorre reversivelmente através da carnitina O-acetiltransferase (CAT, EC:2.3.1.7). O transporte de acetil-CoA entre os

compartimentos intracelulares é mediado por carnitinas acetiltransferases que reversivelmente ligam unidades acetil à molécula carreadora de carnitina. Acetil-CoA é o metabólito central que é produto e substrato de muitas vias metabólicas que participam do metabolismo de carbono (379). Durante a síntese de fontes de carbono, como ácidos graxos, etanol ou acetato, acetil-CoA é produzida em diferentes locais na célula, que necessitam o transporte de unidades de acetil entre diferentes compartimentos (380). Acetil-CoA não consegue atravessar a membrana das organelas sem carnitina (381). O efeito nocivo da inibição de CATs foi descrito em diversos fungos (382–384). Justamente, Jaudzems et al. (385) demonstraram que o efeito cardioprotetor do mildronato (3-(2,2,2-trimetil-hidrazínio)-propionato) é capaz de inibir a CAT através da ligação da enzima competitivamente com acetil-CoA.

5.2.1.4 3-hidroxi-antranilato 3,4-dioxigenase

A enzima hidroxiantranilato 3,5 dioxigenase (EC:1.13.11.6) atua catalisando a abertura do anel de 3-hidroxi-antranilato no passo enzimático final na via biosintética do triptofano para quinolinato. A biosíntese do quinolinato a partir do triptofano possui importância chave, uma vez que o quinolinato é o precursor do anel de piridina dos cofatores de nicotinamida (386). Além disso, a enzima 3-hidroxi-antranilato 3,4-dioxigenase atua na via da quinurenina, que é considerada a principal rota para o metabolismo de triptofano para gerar quinolinato, o que justifica a enzima como um alvo enzimático em potencial para o desenvolvimento de um inibidor (387). Embora a enzima 3-hidroxi-antranilato 3,4-dioxigenase não seja documentada na literatura para *F. graminearum* e para *F. oxysporum*, esta pode ser considerada de interesse para o desenvolvimento de um fungicida, uma vez que esta é a última enzima envolvida na

biossíntese de quinolinato e ao mesmo tempo se encontra na entrada da via para o metabolismo de nicotinamida.

5.2.1.5 Ureidoglicolato liase

A enzima ureidoglicolato liase (EC:4.3.2.3) é responsável por catalisar a degradação do ureidoglicolato em glicoxalato e ureia; esta enzima se encontra atuando no passo final da via catabólica levando purinas para uréia (388). McIninch et al. (388) apresentaram um inibidor para a enzima ureidoglicolato liase designado para mimetizar as espécies transitórias produzidas durante a sua catálise; o inibidor foi denominado ácido 2,4-dioxo-4-fenilbutanóico e foi considerado um dos mais potentes inibidores competitivos com um K_i de 2,2 nM.

5.2.1.6 Holocitocromo-c sintase

A enzima holocitocromo-c sintase (HCCS, citocromo c heme liase, heme liase, EC:4.4.1.17) catalisa a importação do apocitocromo para dentro da mitocôndrias de animais, fungos e protozoários através da ligação covalente do cofator heme ao citocromo c, que é um componente essencial da cadeia de transporte de elétrons das mitocôndrias (389,390). A formação de holocitocromo c pode ser inibida pela deuterohemina, que é considerada um análogo do substrato de HCCS (391).

5.2.1.7 Esfingomielinase neutra

As enzimas esfingomielinases neutra (N-SMases, esfingomielina fosfodiesterase 2, EC:3.1.4.12) são consideradas as principais candidatas para mediar o estresse induzido pela produção de ceramida. A atividade N-SMase foi descrita, caracterizada e clonada de células bacteriana, leveduras e células de mamíferos. Embora o nível de identidade entre estas enzimas seja baixo, os resíduos chaves que estão envolvidos na ligação de íons metálicos e na catálise são conservados (392).

Em leveduras, a síntese de esfingolípídeos é vital para o crescimento e a inibição de N-SMases induz a morte celular (393). Por exemplo, a australifungina é um inibidor que é capaz de bloquear a síntese de ceramida em leveduras (394), entretanto este inibidor não é comercialmente disponível. Nagiec et al. (393) demonstraram que o antifúngico aureobasidina A foi capaz de inibir a enzima inositol fosforilceramida em leveduras ($IC_{50} = 0,2 \text{ nM}$).

5.2.2 Enzimas Auxiliares

5.2.2.12-metilcitrato desidratase

As espécies de *Fusarium* possuem genes que codificam enzimas putativas do ciclo do metilcitrato, no entanto, não houve evidência experimental de que o ciclo do metilcitrato seja funcional nessas espécies (395). A enzima 2-metilcitrato desidratase, EC:4.2.1.79 é responsável pela conversão de 2-metilcitrato em 2-metil-cis-aconitato no ciclo do metilcitrato, que é responsável pela produção de metabolitos secundários altamente tóxicos durante a infecção da planta (395–398).

O acúmulo de propanoil-CoA exerce um efeito tóxico inibindo várias enzimas do metabolismo primário (399). Para evitar este acúmulo, a maioria dos fungos utiliza o ciclo do metilcitrato para a degradação eficiente de propanoil-CoA em piruvato. A inibição desta enzima levaria a um acúmulo de 2-metilcitrato, embora haja uma rota alternativa em *F. oxysporum*, como pode ser deduzido a partir da análise do mapa KEGG fgr00640.

5.2.2.2 Indoleamina 2,3-dioxigenase

A enzima indoleamina 2,3 dioxigenase (IDO, EC:1.13.11.52) catalisa o primeiro passo no catabolismo do triptofano (L-Trp) através da via da quinurenina (400), e seu papel principal é fornecer dinucleótido de nicotinamida e adenina (NAD⁺) através da via de quinurenina (401). Como é uma enzima envolvida em patologias humanas (402–406), alguns inibidores já foram descritos na literatura para IDO (405,407,408). Jang et al. (409,410) isolaram benzomalvina E de *Penicillium* sp. FN070315 e thielavinas de *Coniochaeta* sp. 10F058 e descreveram estes metabólitos como inibidores de IDO.

5.2.2.3 Fosfatidiletanolamina N-metiltransferase

A enzima fosfatidiletanolamina N-metiltransferase (EC:2.1.1.17) catalisa a transferência de um ou mais grupos metil da S-adenosil-L-metionina. Nos fungos, a via de metilação é utilizada como a principal via de biossíntese de fosfatidilcolina (411,412). Uma vez que a via CDP-colina é uma via auxiliar para gerar fosfatidilcolina, que é essencial para os fungos, a inibição dessa via também se torna necessária (413).

Robson et al. (414) demonstraram que o inibidor não competitivo hemicholinium-3 altera a morfologia das hifas na cepa *F. graminearum* A3/5, o que sugere que os inibidores dessa via devem ser considerados em mais detalhes.

5.2.2.4 L-metionina (R)-S-oxidase redutase

A enzima L-metionina (R)-S-óxido redutase (EC:1.8.4.14) catalisa a síntese de L-Metionina (L-Met), que é um aminoácido essencial. Embora a síntese de L-Met seja primordial para o crescimento e desenvolvimento de fungos, a inibição da enzima L-metionina (R)-S-oxidase redutase não seria suficiente para causar a morte do patógeno. Outra rota alternativa para a produção de L-Met é encontrada em *Fusarium* através da via L-Homocisteína, que possui duas enzimas que são capazes de gerar L-Met, homocisteína S-metiltransferase (EC:2.1.1.10) e 5-metiltetrahydropterotrilglutamato-homocisteína-S-metiltransferase (EC:2.1.1.14).

5.2.2.5 ATP adenililtransferase

A enzima ATP adenililtransferase (EC:2.7.7.53) é responsável por catalisar a degradação de Ap4A em ATP e ADP na presença de fosfato inorgânico (415). Entretanto, de acordo com Plateau et al. (415), a inibição de apa1 e/ou apa2 não afetaria a viabilidade celular.

5.3 Validação das Enzimas Específicas por Análise de balanço de fluxo –

FBA

5.3.1 Análise de balanço de fluxo da reconstrução metabólica de *F. oxysporum*

A reconstrução metabólica na escala genômica é uma ferramenta disponível para a modelagem metabólica que pode atuar tanto na descoberta de novos alvos quanto ao desenvolvimento de medicamentos e na compreensão de doenças humanas (21,416,417). Os métodos computacionais adequados para a reconstrução metabólica devem ser capazes de produzir modelos que requerem apenas uma quantidade mínima de intervenção humana e que também possam prever com precisão os fenótipos metabólicos, uma vez que o delúgio de dados produzidos por meio do sequenciamento do genoma está criando gargalos nas análises metabólicas (418).

A utilidade do FBA tem sido cada vez mais reconhecida por sua simplicidade e extensibilidade (419,420), uma vez que requer apenas informações sobre a estequiometria da reação metabólica e equilíbrio de massa em torno de metabolitos sob estado pseudo-estável (421). Além disso, o FBA foi aplicado com sucesso a uma multiplicidade de espécies como *Chlamydomonas reinhardtii* (422), *E. coli* (423,424), *Neisseria meningitidis* (425), *S. cerevisiae* iND750 (426), *Staphylococcus aureus* N315 (427), *Streptococcus thermophilus* LMG18311 (428) para modelar seus metabolismos celulares.

Determinar a composição da biomassa é o primeiro passo para a análise de fluxo. Ela é uma componente chave para prever o metabolismo celular, através da distribuição do fluxo em modelos metabólicos baseados em restrição usando FBA, pois

os fluxos intracelulares são dependentes dos fluxos que contribuem para a síntese de biomassa (429,430). E embora tenham ocorrido progressos nas tecnologias analíticas, muitas vezes é difícil obter uma composição detalhada de todos os principais componentes da biomassa experimentalmente.

Evidências experimentais indicam que a composição da biomassa varia entre espécies, tipos de células e condições fisiológicas (431,432). No entanto, devido à falta de informações experimentais específicas de organismo e/ou condições específicas, as composições de biomassa usadas nos modelos metabólicos em escala genômica de plantas são frequentemente coletadas de diversos tipos de medições, experimentos, grupos de pesquisa e até diferentes tipos de células e espécies de plantas (433).

Em plantas, pouco se sabe sobre o impacto da variação da composição da biomassa na previsão de fluxo em modelos utilizando FBA, cujo metabolismo é muito versátil e complexo devido à presença de múltiplos compartimentos subcelulares. Além disso, os modelos metabólicos publicados de plantas diferem em tamanho e complexidade (434).

Um estudo realizado por Yuan et al., (2016) (434) investigou o efeito da composição da biomassa no estado estacionário em três modelos metabólicos de *Arabidopsis* (435), AraGEM (300) e AraCore (436). Foi observado que no modelo de Poolman et al., (2009) (435) a parede celular compreendia mais da metade da biomassa celular. Em contraste, as células continham maiores quantidades de proteína no modelo AraGEM e AraCore. Evidências experimentais indicam que as distribuições dos componentes da biomassa são muito específicas do tecido (437,438). Um resultado esperado é a semelhança entre a composição das células no modelo AraGEM e AraCore, uma vez que ambos modelos representam células foliares fotossintéticas. E embora

existam algumas semelhanças em termos de proporções de parede celular, carboidrato e proteína, também há grandes diferenças. Por exemplo, a proporção de lipídios na biomassa AraCore é muito maior do que na biomassa AraGEM (18,4 vs. 1,1%). Isso pode ser explicado pelo fato de que os dados de biomassa no modelo AraGEM são coletados de diferentes organismos (434).

Em contrapartida, observamos que os componentes de biomassa de *F. oxysporum* são em sua maioria compostos de aminoácidos (45%), seguidos de nucleotídeo (25%) de carboidratos (8%), lipídeos (4%) e componentes de menor importância (18%). Estes resultados são provenientes do modelo disponível na época (293) deste estudo (Biomodel MODEL1302010014) e inspirado nos dados de levedura; ele correspondem basicamente ao que se obteve experimentalmente analisando o valor nutritivo dos cogumelos (439). Considerando a produção de aminoácidos, uma análise de quatro cepas fúngicas miceliais e uma levedura em culturas colhidas nas fases de crescimento médio e no início do crescimento estacionário, Christias et al. (1975) (440) mostraram que *F. oxysporum* continha aproximadamente 325 mg aminoácidos por g de micélia (32,5%) em cultura o que é levemente abaixo do valor reportado no modelo.

Contudo, diferenças importantes podem ser observadas com os valores dos metabólitos registrados ao nível experimental no caso de *F. oxysporum*. Por exemplo, em um estudo promovido por (441), a biomassa do fungo estudado aumentou a uma taxa lenta até cerca de 48h, mas depois aumentou a uma taxa exponencial. A natureza e a disponibilidade de substratos de crescimento e carboidratos são cruciais para prever a taxa de crescimento e a multiplicação de fungos em culturas de laboratório. Por exemplo, ao analisar diferentes fontes de carbono (meios sintéticos ou caules de sorgo doce) e nitrogênio, bem como a relação carbono/nitrogênio, usadas para o cultivo de *F. oxysporum*, foi possível obter concentração de lipídios de 4,4 g/L com o conteúdo de

lipídeos em 42,6% (w/w). Além disso, o fungo foi capaz de crescer em caules de sorgo sacarificado sem a necessidade de suplementação externa de nitrogênio. A remoção das partículas de caules de sorgo doce permitiu a produção de lipídios durante o cultivo em sólidos iniciais aumentados em até 16 (w/w), o que resultou em uma produção lipídica de 3,81 g/L (442).

Feofilova et al. (2015) (443) ao avaliar a composição lipídica da parede celular, células micélicas inteiras e de células dormentes de fungos mucoráceos e ascomicetos evidenciaram que células miceliais inteiras continham mais lipídios que a parede celular. O valor do conteúdo lipídico nas células miceliais inteiras nas três espécies de fungos variou de 10% a 50% do valor total da biomassa. Ao contrário, em células fúngicas no estado de dormência exógena (esporos), a parede celular obteve um maior conteúdo lipídico e variou de 7% a 30% do valor total da biomassa (443).

Naturalmente, estes valores de conteúdo lipídico são o resultado de um processo de acumulação em vacúolos que pode ser específico de um determinado fungo. Estes valores não corroboram o fluxo metabólico de biosíntese do modelo que ele é apenas citosólico e provavelmente relativamente conservado em fungos.

Contudo, a modelagem FBA em fungo é uma atividade muito nova no campo da ciência e a significância dos resultados do modelo analisado aqui, a pesar de ter produzidos inferências de alvos valiosas, têm que ser considerados com cuidado.

Em relação às enzimas que foram identificadas como específicas de *F. oxysporum* considerando cinco espécies de plantas hospedeiras e que foram inseridas na reconstrução do fungo, estas não puderam ser consideradas críticas para a sobrevivência do *F. oxysporum*. Porém, é claro que essa suposição deve ser confirmada por experimentação de bancada, pois, de fato, as seis enzimas são principalmente

específicas. E de acordo com os mapas metabólicos do KEGG, essas enzimas possuem papel de centralidade nesses mapas. Devido a essa representação simplista dos mapas metabólicos, a avaliação da centralidade dos mapas do KEGG é provavelmente muito tendenciosa.

5.3.2 Genes e atividades de enzimas inseridas na Reconstrução metabólica de *F. oxysporum*

Como mencionado na seção 4.3.1, foi possível atualizar a reconstrução metabólica de *F. oxysporum* realizada por Pitkänen et al. (2014) (293) através da inclusão de 6 novos genes e 6 novas atividades enzimáticas. Destes, dois genes representados por duas atividades enzimáticas (F9F7R0, ureidoglicolato liase – EC: 4.3.2.3 e F9FE93, m7GpppX difosfatase – EC: 3.6.1.59) foram considerados chave para o controle de *Fusarium* spp.; e um gene representado por uma atividade enzimática (F9FYW8, ATP adenililtransferase – EC: 2.7.7.53) foi considerado uma enzima auxiliar. Estes tiveram seus papéis fisiológicos discutidos na seção 5.2.1, bem como a existência de inibidores a fim de justificar a sua inibição. De modo que estaremos discutindo o papel fisiológico de apenas três genes incluídos na reconstrução metabólica do fungo representados por três atividades enzimáticas (F9FX31, 25S rRNA M(1)2142 metiltransferase – EC:2.1.1.86; F9FL57, carboxi-cis, cis-muconato ciclase – EC:5.5.1.5 e F9FPS4, aconitato descarboxilase – EC:4.1.1.6) a seguir.

5.3.2.1 25S rRNA M(1)2142 – metiltransferase

A metilação é uma modificação pós-transcricional predominante encontrada no rRNA (444). A metilação da base promove empilhamento de base aumentando a

hidrofobicidade e expandindo a polarizabilidade (445). A enzima 25S rRNA M(1)2142 metiltransferase também conhecida como Bmt2 está envolvida na modificação de base do 25S rRNA na posição 2142. A enzima Bmt2 é a segunda enzima m¹A metiltransferase envolvida na modificação de base da hélice 65 da 25S rRNA em *S. cerevisiae*. A falta de modificação fornece peróxido de hidrogênio e sensibilidade à anisomicina, juntamente com um leve defeito na união das subunidades (446).

A observações feitas através de modificação de base de tRNA e m¹A 2142 indicam que as vias de biossíntese destas modificações podem estar envolvidas na resposta celular no sentido de estresse oxidativo estimulada por peróxido de hidrogênio. No entanto, estudos futuros são necessários para elucidar o papel específico dessas modificações no estresse oxidativo (447).

5.3.2.2 Carboxi-cis,cis-muconato ciclase

A enzima carboxi-cis,cis-muconato ciclase pertence à família das isomerases, especificamente à classe das liases intramoleculares. É uma enzima que participa na degradação do benzoato via hidroxilação.

No fungo *F. oxysporum*, o mutante que apresenta a deleção desta enzima fica inviabilizado de catabolizar compostos fenólicos conhecidos por serem degradados através da via do β - ceto adipato. E como consequência, o mutante é prejudicado na invasão radicular e se torna não patogênico, apesar de apresentar colonização radicular superficial normal (448).

A deleção desta enzima não acarreta a morte do fungo, e sim, a inviabilidade patogênica deste. Entretanto, por não possuir um mapa metabólico definido que atende

os requisitos para esta ser classificada como uma enzima auxiliar (tópico 5.2.2), esta enzima somente foi classificada com uma enzima importante já que identificada pela metodologia de busca por enzimas específicas e foi inserida na reconstrução metabólica de *F. oxysporum*.

5.3.2.3 Aconitato descarboxilase

A enzima aconitato descarboxilase participa da via de biossíntese do ácido itacônico (449,450). A função da enzima aconitato descarboxilase foi descrita no fungo filamentosos *Aspergillus terreus* (451). A enzima catalisa a formação do ácido itacônico pela descarboxilação do cis-acetonato para o ácido itacônico (452).

A aconitato descarboxilase (aconitase) é a primeira reação de descarboxilação experimentalmente demonstrada do ácido itacônico, levando à formação do grupo metileno: cis-acetonato \rightleftharpoons itaconato + CO₂, onde o ácido acético é descarboxilado com a formação de uma dupla ligação entre o carbono 2 (C2) e o C3 do ácido cítrico para produzir o itaconato. O itaconato é um metabólito antibacteriano (453) que é produzido pela fermentação de fungos filamentosos na presença de oligossacarídeos (454).

O mapa do MetaCyc mostra a rota de biossíntese do itaconato a partir do piruvato (<https://biocyc.org/META/NEW-IMAGE?type=NIL&object=PWY-5750>). Após a remoção de uma molécula de água do citrato pela aconitase, que ocorre na mitocôndria, o cis-acetonato é movido para o citosol, onde é convertido em itaconato pela aconitato descarboxilase (455). Contudo, este mapa não atende os requisitos para que esta enzima pudesse ser classificada como uma enzima principal ou auxiliar (tópico 5.2.2). Então, na

dúvida, inserimos ela na reconstrução de *F. oxysporum* como fizemos para a carboxi-cis,cis-muconato ciclase.

5.3.3 Modelagem tridimensional das duas enzimas análogas consideradas críticas

Os dois pares de enzimas F9F4G5 (*F. oxysporum*) - Q8GWP5 (*A. thaliana*) e F9FSB6 (*F. oxysporum*) - Q9SZE9 (*A. thaliana*) foram encontrados de forma análoga e em um deles F9F4G5 - Q8GWP5, ambas as enzimas podiam ser comparadas, pois os modelos 3D estavam disponíveis no PDB.

A modelagem 3D permitiu confirmar o caráter análogo de pelo menos um par de enzimas já que o potencial eletrostático do domínio catalítico da enzima de *F. oxysporum* (F9F4G5) é principalmente eletropositivo quando o potencial eletrostático do domínio da enzima de *A. thaliana* (Q8GWP5) é principalmente eletronegativo. Por consequência, um inibidor carregado negativamente poderá só interagir com uma e vice-versa o que demonstra o caráter seletivo que um inibidor deste tipo pode possuir e por consequência o potencial específico das enzimas análogas no tratamento terapêutico.

A enzima análoga do fungo, F9F4G5, é crítica segundo a classificação do FBA e a sua inibição teria potencial de alterar de forma significativa o próprio crescimento do *F. oxysporum*. A enzima pseudouridina sintase, EC: 3.5.4.26, é responsável pela modificação pós-transcricional mais frequente de RNAs celulares. Ela catalisa a isomerização específica de resíduos de uridina que já fazem parte de uma cadeia de RNA e parece empregar informações de sequência e estruturas 3D para atingir a sua especificidade (456).

King & Lu (457) avaliaram a significância biológica das pseudouridinas sintase eucarióticas usando a levedura *S. cerevisiae* e observaram que as cepas defetivas apresentaram alteração diferencial nas propriedades de crescimento do organismo, incluindo, em alguns casos, crescimento aumentado e/ou curvas de crescimento bifásico não observadas em cepas de tipo selvagem em condições equivalentes. Os resultados demonstram que a inibição da pseudouridina sintase pode ter um efeito qualitativo significativo no crescimento.

Além disso, em *T. gondii* o gene da pseudouridina sintase é crítico na diferenciação celular entre as duas formas assexuadas: taquizoítos e bradioítos (458). Em humanos, a mutação nos genes responsáveis pela síntese de pseudouridina sintase é responsável por doenças (miopatia mitocondrial e anemia sideroblástica - MLASA; disqueratose congênita) (459,460).

A enzima de *F. oxysporum* que não pôde ser modelada por possuir baixa identidade com os moldes disponíveis no PDB, F9FSB6 - EC; 3.5.1.49, é uma formamidase. A enzima formamidase medeia a hidrólise altamente específica de formamida para produzir formato e amônia (461). Em *Aspergillus* spp., a formamida, serve apenas como fonte de nitrogênio, pois o formiato tem apenas um átomo de carbono e, portanto, não é uma fonte de carbono(462).

Nos fungos filamentosos, as atividades de muitas enzimas catabólicas de nitrogênio, incluindo as amidases, mostraram-se altamente reguladas em resposta aos níveis de nitrogênio na célula através da repressão do metabólito de nitrogênio, o que resulta na utilização preferencial das fontes de nitrogênio de amônia e L-glutamina (463). Para fungos fitopatógenos foi proposto que a disponibilidade de nitrogênio durante o crescimento da planta pode ser um fator limitante (464–468). Assim, de

acordo com o estilo de vida particular destes organismos, os desafios nutricionais podem exigir várias adaptações. Os fungos podem utilizar uma ampla gama de fontes de nitrogênio. A amônia e a glutamina são preferidas, mas quando essas fontes primárias estão ausentes ou estão presentes em baixas concentrações, outras fontes secundárias podem ser usadas (463).

para os alvos da rede metabólica em continuação de um trabalho pioneiro iniciado no Instituto Oswaldo Cruz (Laboratório de Genômica Funcional e Bioinformática) (37).

7. CONCLUSÃO

Como conclusões do trabalho podemos citar:

- Com a metodologia de busca de enzimas específicas desenvolvida conseguimos identificar 40 enzimas estritamente específicas de *L. major* no seu sistema parasitário com *H. sapiens*.
- Das 40 enzimas estritamente específicas de *L. major*, 32 enzimas foram descartadas por apresentarem vias alternativas no metabolismo do parasito. Neste estudo, somente foram sugeridas como alvos para o desenvolvimento de fármacos enzimas essenciais para *L. major* que possuíram vias únicas no metabolismo do parasito.
- As enzimas esterol 24-C-metiltransferase, piruvato fosfato diquinase, tripanotiona sintetase e RNA-edição ligase foram sugeridas como sendo enzimas essenciais para *L. major* que podem servir como alvos para o desenvolvimento de fármacos.
- Foi possível implementar a metodologia de busca por enzimas específicas em outro sistema hospedeiro-parasito, no caso, planta-fungo: *A. thaliana* e *F. oxysporum*; *B. rapa* e *F. oxysporum*; *G. max* e *F. oxysporum*; *J. curcas* e *F. oxysporum*; *R. communis* e *F. oxysporum*.
- Durante a implementação da metodologia de busca por enzimas específicas no sistema planta-fungo foram identificadas 31 enzimas específicas para *F. graminearum* e 28 enzimas específicas para *F. oxysporum* em relação às cinco espécies de plantas estudadas.

- Com o estudo de enzimas específicas de *F. oxysporum*, foi possível inserir seis genes, F9FX31, F9FYW8, F9FE93, F9FPS4, F9F7R0, F9FL57, que não estavam presentes na reconstrução metabólica do *F. oxysporum*.
- A modelagem por FBA permitiu identificar três enzimas críticas de *F. oxysporum*, B0SYU4, F9F4G5 e F9FSB6, para a produção de biomassa, cuja inibição teria potencialidade de desarticular a rede metabólica do fungo *in vivo*.
- As enzimas críticas para a produção de biomassa pelo fungo, F9F4G5 (EC: 3.5.4.26) e F9FSB6 (EC: 3.5.1.49), foram consideradas análogas por possuírem uma contraparte na planta *A. thaliana*, Q8GWP5 (EC: 3.5.4.26) e Q9SZE9 (EC: 3.5.1.49).
- Foi possível modelar a estrutura 3D de somente o par de uma das enzimas consideradas críticas para a produção de biomassa pelo fungo F9F4G5 (EC: 3.5.4.26), de *F. oxysporum* com seu respectivo par Q8GWP5 (EC: 3.5.4.26).
- O gene B0SYU4, apesar não ter homólogos próximos no genoma de *F. oxysporum*, foi incluído na reconstrução metabólica deste fungo por ser considerado crítico para a produção de biomassa (*gap-filling*).
- A enzima pseudouridina sintase, F9F4G5 - EC: 3.5.4.26, pode ser considerada um alvo apropriado à inibição por RNA interferência produzido por hospedeiro transformados para adquirir esta competência já que envolvido no crescimento do fungo e regulação de processos biológicos essenciais.

■ A enzima formamidase, F9FSB6 - EC; 3.5.1.49, pode ser considerada um alvo secundário para desenvolvimento de um fungicida, pois não está envolvida com uma função vital para sobrevivência do fungo. Apesar desta enzima estar envolvida com a hidrólise de formamida para produzir amônia, o fungo consegue utilizar outras fontes secundárias quando a amônia está ausente ou presente em baixa concentração.

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. Avorn J. The \$2.6 billion pill—methodologic and policy considerations. *N Engl J Med.* 2015;372(20):1877–9.
2. Tufts Center for the Study of Drug Development. How the Tufts Center for the Study of Drug Development pegged the cost of a new drug at \$2.6 billion. Boston: Tufts Center for the Study of Drug Development.
3. Altshuler J, Flanagan A, Guy P, Steiner M, Tollman P. A Revolution in R&D: How Genomics and Genetics are Transforming the Biopharmaceutical Industry. Bost Consult Gr. 2001;
4. Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol.* 2001;19(1):45–50.
5. Oliver SG. From DNA sequence to biological function. *Nature.* 1996;379(6566):597–600.
6. Cornish-Bowden A, Cárdenas ML. Complex networks of interactions connect genes to phenotypes. *Trends Biochem Sci.* 2001;26(8):463–5.
7. Pennisi E. ENCODE project writes eulogy for junk DNA. *Science* (80-). 2012;337(6099):1159–61.
8. Chen X-P, Du G-H. Target validation: A door to drug discovery. *Drug Discov Ther.* 2007;1(1):23–9.
9. Bull SC, Doig AJ. Properties of Protein Drug Target Classes. *PLoS One.*

- 2015;10(3):e0117955.
10. Imming P, Sinning C, Meyer A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov.* outubro de 2006;5(10):821–34.
 11. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 2013;42(D1):D1091–7.
 12. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov.* 2017;16(1):19–34.
 13. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov.* 2002;1(9):727–30.
 14. Xu H, Xu H, Lin M, Wang W, Li Z, Huang J, et al. Learning the drug target-likeness of a protein. *Proteomics.* 2007;7(23):4255–63.
 15. Seringhaus M, Paccanaro A, Borneman A, Snyder M, Gerstein M. Predicting essential genes in fungal genomes. *Genome Res.* 2006;16(9):1126–35.
 16. Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee PW-N. Metabolic control analysis in drug discovery and disease. *Nat Biotechnol.* 2002;20(3):243–9.
 17. Haanstra JR, Gerding A, Dolga AM, Sorgdrager FJH, Buist-Homan M, Du Toit F, et al. Targeting pathogen metabolism without collateral damage to the host. *Sci Rep.* 2017;7:40406.

18. Galperin MY, Walker DR, Koonin E V. Analogous enzymes: Independent inventions in enzyme evolution. *Genome Res.* 1998;8(8):779–90.
19. Fitch WM. Distinguishing homologous from analogous proteins. *Syst Zool.* 1970;19(2):99–113.
20. Ruppin E, Papin JA, De Figueiredo LF, Schuster S. Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr Opin Biotechnol.* 2010;21(4):502–10.
21. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol.* 2011;7(1):501.
22. Sprinzak E, Margalit H. Correlated sequence-signatures as markers of protein-protein interaction. *J Mol Biol.* 2001;311(4):681–92.
23. Wan KK, Park J, Suh JK. Large scale statistical prediction of protein-protein interaction by potentially interacting domain (PID) pair. *Genome Informatics.* 2002;13:42–50.
24. Ng S-K, Zhang Z, Tan S-H. Integrative approach for computationally inferring protein domain interactions. In: *Proceedings of the 2003 ACM symposium on Applied computing.* 2003. p. 115–21.
25. Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han J-DJ, et al. Annotation transfer between genomes: protein–protein interologs and protein–DNA regulogs. *Genome Res.* 2004;14(6):1107–18.
26. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data.

- Science (80-). 2003;302(5644):449–53.
27. Zhang L V, Wong SL, King OD, Roth FP. Predicting co-complexed protein pairs using genomic and proteomic data integration. *BMC Bioinformatics*. 2004;5(1):38.
 28. Qi Y, Bar-Joseph Z, Klein-Seetharaman J. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins Struct Funct Bioinforma*. 2006;63(3):490–500.
 29. Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein J Org Chem*. 2016;12:2694.
 30. Capriles PVSZ, Baptista LPR, Guedes IA, Guimarães ACR, Custódio FL, Alves-Ferreira M, et al. Structural modeling and docking studies of ribose 5-phosphate isomerase from *Leishmania major* and *Homo sapiens*: A comparative analysis for Leishmaniasis treatment. *J Mol Graph Model*. fevereiro de 2015;55:134–47.
 31. Kanehisa M, Bork P. Bioinformatics in the post-sequence era. *Nat Genet*. 2003;33.
 32. Gaasterland T, Sensen CW. Fully automated genome analysis that reflects user needs and preferences. A detailed introduction to the MAGPIE system architecture. *Biochimie*. 1996;78(5):302–10.
 33. Fleischmann W, Møller S, Gateau A, Apweiler R. A novel method for automatic functional annotation of proteins. *Bioinformatics*. 1999;15(3):228–33.
 34. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med*. 2001;40(4):346–58.

35. Benner SA, Gaucher EA. Evolution, language and analogy in functional genomics. *TRENDS Genet.* 2001;17(7):414–8.
36. Colebatch G, Trevaskis B, Udvardi M. Functional genomics: tools of the trade. *New Phytol.* 2002;153(1):27–36.
37. Otto TD, Guimarães ACR, Degraeve WM, de Miranda AB. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics.* 2008;9:544.
38. Kanehisa M. Post-genome informatics. 2000.
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
40. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39(suppl_2):W29–37.
41. Lindoso JAL, Lindoso AABP. Neglected tropical diseases in Brazil. *Rev Inst Med Trop Sao Paulo.* 2009;51(5):247–53.
42. World Health Organization. Leishmaniasis: burden of disease. 2009.
43. Werneck GL. Leishmaniose visceral no Brasil: fundamentos e preocupações em relação ao controle de reservatórios. *Rev Saude Publica.* 2014;48(5):851–6.
44. Waldman EA, Sato APS. Path of infectious diseases in Brazil in the last 50 years: an ongoing challenge. *Rev Saude Publica.* 2016;50.
45. McCall L-I, McKerrow JH. Determinants of disease phenotype in trypanosomatid parasites. *Trends Parasitol.* 2014;30(7):342–9.
46. Kaur G, Rajput B. Comparative analysis of the omics technologies used to study

- antimonial, amphotericin B, and pentamidine resistance in leishmania. *J Parasitol Res.* 2014;2014.
47. Pawar H, Renuse S, Khobragade SN, Chavan S, Sathe G, Kumar P, et al. Neglected tropical diseases and omics science: proteogenomics analysis of the promastigote stage of *Leishmania major* parasite. *Omi a J Integr Biol.* 2014;18(8):499–512.
 48. Cantacessi C, Dantas-Torres F, Nolan MJ, Otranto D. The past, present, and future of *Leishmania* genomics and transcriptomics. *Trends Parasitol.* 2015;31(3):100–8.
 49. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860–921.
 50. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science (80-).* 2001;291(5507):1304–51.
 51. Abdellah Z, Ahmadi A, Ahmed S, Aimable M, Ainscough R, Almeida J. International human genome sequencing consortium. *Nature.* 2004;409:860–921.
 52. Muñoz J, Heck AJR. From the human genome to the human proteome. *Angew Chemie Int Ed.* 2014;53(41):10864–6.
 53. Bohlander SK. ABCs of genomics. *ASH Educ Progr B.* 2013;2013(1):316–23.
 54. Simonti CN, Capra JA. The evolution of the human genome. *Curr Opin Genet Dev.* 2015;35:9–15.

55. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 2012;22(9):1760–74.
56. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621–8.
57. Coffey AJ, Kokocinski F, Calafato MS, Scott CE, Palta P, Drury E, et al. The GENCODE exome: sequencing the complete human exome. *Eur J Hum Genet.* 2011;19(7):827–31.
58. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One.* 2012;7(5):e35671.
59. Akhouni M, Kuhls K, Cannet A, Votýpka J, Marty P, Delaunay P, et al. A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Negl Trop Dis.* 2016;10(3):e0004349.
60. Singh S, Dey A, Sivakumar R. Applications of molecular methods for *Leishmania* control. *Expert Rev Mol Diagn.* 2005;5(2):251–65.
61. Chawla B, Madhubala R. Drug targets in *Leishmania*. *J Parasit Dis.* 2010;34(1):1–13.
62. Sundar S, Chakravarty J. An update on pharmacotherapy for leishmaniasis. *Expert Opin Pharmacother.* 2015;16(2):237–52.
63. Moen MD, Lyseng-Williamson KA, Scott LJ. Liposomal amphotericin B. *Drugs.* 2009;69(3):361–92.

64. Wyllie S, Cunningham ML, Fairlamb AH. Dual action of antimonial drugs on thiol redox metabolism in the human pathogen *Leishmania donovani*. *J Biol Chem*. 2004;279(38):39925–32.
65. Baiocco P, Colotti G, Franceschini S, Ilari A. Molecular basis of antimony treatment in leishmaniasis. *J Med Chem*. 2009;52(8):2603–12.
66. Frézard F, Demicheli C, Ribeiro RR. Pentavalent antimonials: new perspectives for old drugs. *Molecules*. 2009;14(7):2317–36.
67. Chawla B, Jhingran A, Panigrahi A, Stuart KD, Madhubala R. Paromomycin affects translation and vesicle-mediated trafficking as revealed by proteomics of paromomycin-susceptible-resistant *Leishmania donovani*. *PLoS One*. 2011;6(10):e26660.
68. Rakotomanga M, Blanc S, Gaudin K, Chaminade P, Loiseau PM. Miltefosine affects lipid metabolism in *Leishmania donovani* promastigotes. *Antimicrob Agents Chemother*. 2007;51(4):1425–30.
69. Verma NK, Dey CS. Possible mechanism of miltefosine-mediated death of *Leishmania donovani*. *Antimicrob Agents Chemother*. 2004;48(8):3010–5.
70. World Health Organization. 2010.
71. Ouellette M, Drummel-Smith J, Papadopoulou B. Leishmaniasis: drugs in the clinic, resistance and new developments. *Drug Resist Updat*. 2004;7(4):257–66.
72. Khan I, Khan M, Umar MN, Oh D-H. Attenuation and production of the amphotericin B-resistant *Leishmania tropica* strain. *Jundishapur J Microbiol*. 2016;9(6).

73. Maspi N, Abdoli A, Ghaffarifar F. Pro-and anti-inflammatory cytokines in cutaneous leishmaniasis: a review. *Pathog Glob Health*. 2016;110(6):247–60.
74. Sacks D, Kamhawi S. Molecular aspects of parasite-vector and vector-host interactions in leishmaniasis. *Annu Rev Microbiol*. 2001;55(1):453–83.
75. Bailey, MS, Lockwood D. Cutaneous leishmaniasis. *Clin dermatology*. 2007;25(2):203–11.
76. Kaye P, Scott P. Leishmaniasis: complexity at the host–pathogen interface. *Nat Rev Microbiol*. 2011;9(8):604–15.
77. Scott P, Novais FO. Cutaneous leishmaniasis: immune responses in protection and pathogenesis. *Nat Rev Immunol*. 2016;
78. Scott P, Natovitz P, Coffman RL, Pearce E, Sher A. Immunoregulation of cutaneous leishmaniasis. T cell lines that transfer protective immunity or exacerbation belong to different T helper subsets and respond to distinct parasite antigens. *J Exp Med*. 1988;168(5):1675–84.
79. Heinzl FP, Sadick MD, Holaday BJ, Coffman RL, Locksley RM. Reciprocal expression of interferon gamma or interleukin 4 during the resolution or progression of murine leishmaniasis. Evidence for expansion of distinct helper T cell subsets. *J Exp Med*. 1989;169(1):59–72.
80. Satoskar A, Bluethmann H, Alexander J. Disruption of the murine interleukin-4 gene inhibits disease progression during *Leishmania mexicana* infection but does not increase control of *Leishmania donovani* infection. *Infect Immun*. 1995;63(12):4894–9.

81. Stamm LM, Räisänen-Sokolowski A, Okano M, Russell ME, David JR, Satoskar AR. Mice with STAT6-targeted gene disruption develop a Th1 response and control cutaneous leishmaniasis. *J Immunol.* 1998;161(11):6180–8.
82. Qi H, Popov V, Soong L. *Leishmania amazonensis*-dendritic cell interactions in vitro and the priming of parasite-specific CD4⁺ T cells in vivo. *J Immunol.* 2001;167(8):4534–42.
83. McMahon-Pratt D, Alexander J. Does the *Leishmania* major paradigm of pathogenesis and protection hold for New World cutaneous leishmaniasis or the visceral disease? *Immunol Rev.* 2004;201(1):206–24.
84. Anderson CF, Mendez S, Sacks DL. Nonhealing infection despite Th1 polarization produced by a strain of *Leishmania major* in C57BL/6 mice. *J Immunol.* 2005;174(5):2934–41.
85. de Moura TR, Novais FO, Oliveira F, Clarêncio J, Noronha A, Barral A, et al. Toward a novel experimental model of infection to study American cutaneous leishmaniasis caused by *Leishmania braziliensis*. *Infect Immun.* 2005;73(9):5827–34.
86. Carvalho EM de, Barral A, Costa JML, Bittencourt A, Marsden P. Clinical and immunopathological aspects of disseminated cutaneous leishmaniasis. *Acta Trop.* 1994;56(4):315–25.
87. Bacellar O, Lessa H, Schriefer A, Machado P, de Jesus AR, Dutra WO, et al. Up-regulation of Th1-type responses in mucosal leishmaniasis patients. *Infect Immun.* 2002;70(12):6734–40.

88. Turk JL, Bryceson ADM. Immunological phenomena in leprosy and related diseases. *Adv Immunol.* 1971;13:209–66.
89. Srivastava S, Shankar P, Mishra J, Singh S. Possibilities and challenges for developing a successful vaccine for leishmaniasis. *Parasit Vectors.* 2016;9(1):277.
90. Palatnik-de-Sousa CB. Vaccines for leishmaniasis in the fore coming 25 years. *Vaccine.* 2008;26(14):1709–24.
91. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *genesis.* 2015;53(8):474–85.
92. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 2011;40(D1):D1202–10.
93. Somerville C, Koornneef M. A fortunate choice: the history of Arabidopsis as a model plant. *Nat Rev Genet.* 2002;3(11):883–9.
94. Koornneef M, Meinke D. The development of Arabidopsis as a model plant. *Plant J.* 2010;61(6):909–21.
95. Strange RN, Scott PR. Plant disease: a threat to global food security. *Annu Rev Phytopathol.* 2005;43.
96. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, et al. Emerging fungal threats to animal, plant and ecosystem health. *Nature.* 2012;484(7393):186–94.

97. Beckman CH. The nature of wilt diseases of plants. 1987.
98. Boutati EI, Anaissie EJ. Fusarium, a Significant Emerging Pathogen in Patients With Hematologic Malignancy: Ten Years⁹ Experience at a Cancer Center and Implications for Management. *Blood*. 1997;90(3):999–1008.
99. Odds FC, Van Gerven F, Espinel-Ingroff A, Bartlett MS, Ghannoum MA, Lancaster M V, et al. Evaluation of possible correlations between antifungal susceptibilities of filamentous fungi in vitro and antifungal treatment outcomes in animal infection models. *Antimicrob Agents Chemother*. 1998;42(2):282–8.
100. Vartivarian SE, Anaissie EJ, Bodey GP. Emerging fungal pathogens in immunocompromised patients: classification, diagnosis, and management. *Clin Infect Dis*. 1993;17(Supplement_2):S487–91.
101. Pontion J, Ruckehl R, Clemons K V, Coleman DC, Grillot R, Guarro J. Emerging pathogens. *Med Mycol*. 2000;38:225–36.
102. Nucci M, Anaissie E. Cutaneous infection by *Fusarium* species in healthy and immunocompromised hosts: implications for diagnosis and management. *Clin Infect Dis*. 2002;35(8):909–20.
103. Weinstein RA, Fridkin SK. The changing face of fungal infections in health care settings. *Clin Infect Dis*. 2005;41(10):1455–60.
104. Nucci M, Anaissie E. *Fusarium* infections in immunocompromised patients. *Clin Microbiol Rev*. 2007;20(4):695–704.
105. Nucci M, Marr KA, Queiroz-Telles F, Martins CA, Trabasso P, Costa S, et al. *Fusarium* infection in hematopoietic stem cell transplant recipients. *Clin Infect*

- Dis. 2004;38(9):1237–42.
106. Fleming R V, Walsh TJ, Anaissie EJ. Emerging and less common fungal pathogens. *Infect Dis Clin North Am.* 2002;16(4):915–33.
 107. Marr KA, Carter RA, Crippa F, Wald A, Corey L. Epidemiology and outcome of mould infections in hematopoietic stem cell transplant recipients. *Clin Infect Dis.* 2002;34(7):909–17.
 108. Nucci M, Anaissie EJ, Queiroz-Telles F, Martins CA, Trabasso P, Solza C, et al. Outcome predictors of 84 patients with hematologic malignancies and *Fusarium* infection. *Cancer.* 2003;98(2):315–9.
 109. Litvinov N, da Silva MTN, van der Heijden IM, Graça MG, de Oliveira LM, Fu L, et al. An outbreak of invasive fusariosis in a children's cancer hospital. *Clin Microbiol Infect.* 2015;21(3):268-e1.
 110. Viscoli C, Castagnola E, Moroni C, Garaventa A, Manno G, Savioli C. Infection with *Fusarium* species in two children with neuroblastoma. *Eur J Clin Microbiol Infect Dis.* 1990;9(10):773–6.
 111. Krcmery V, Jesenska Z, Spanik S, Gyarfás J, Nogova J, Botek R, et al. Fungaemia due to *Fusarium* spp. in cancer patients. *J Hosp Infect.* 1997;36(3):223–8.
 112. Vagace J-M, Sanz-Rodriguez C, Casado M-S, Alonso N, Garcia-Dominguez M, de la Llana FG, et al. Resolution of disseminated fusariosis in a child with acute leukemia treated with combined antifungal therapy: a case report. *BMC Infect Dis.* 2007;7(1):40.

113. Morris SK, Allen UD, Gupta S, Richardson SE. Breakthrough filamentous fungal infections in pediatric hematopoietic stem cell transplant and oncology patients receiving caspofungin. *Can J Infect Dis Med Microbiol.* 2012;23(4):179–82.
114. Guarro J. Fusariosis, a complex infection caused by a high diversity of fungal species refractory to treatment. *Eur J Clin Microbiol Infect Dis.* 2013;32(12):1491–500.
115. Schwartz KL, Sheffield H, Richardson SE, Sung L, Morris SK. Invasive fusariosis: a single pediatric center 15-year experience. *J Pediatric Infect Dis Soc.* 2013;4(2):163–70.
116. Wattier RL, Dvorak CC, Hoffman JA, Brozovich AA, Bin-Hussain I, Groll AH, et al. A prospective, international cohort study of invasive mold infections in children. *J Pediatric Infect Dis Soc.* 2014;4(4):313–22.
117. Carlesse F, Amaral A-PC, Gonçalves SS, Xafranski H, Lee M-LM, Zecchin V, et al. Outbreak of *Fusarium oxysporum* infections in children with cancer: an experience with 7 episodes of catheter-related fungemia. *Antimicrob Resist Infect Control.* 2017;6(1):93.
118. Ortoneda M, Guarro J, Madrid MP, Caracuel Z, Roncero MIG, Mayayo E, et al. *Fusarium oxysporum* as a multihost model for the genetic dissection of fungal virulence in plants and mammals. *Infect Immun.* 2004;72(3):1760–6.
119. López-Díaz C, Rahjoo V, Sulyok M, Ghionna V, Martín-Vicente A, Capilla J, et al. Fusaric acid contributes to virulence of *Fusarium oxysporum* on plant and mammalian hosts. *Mol Plant Pathol.* 2017;

120. Mitchell-Olds T. *Arabidopsis thaliana* and its wild relatives: a model system for ecology and evolution. *Trends Ecol Evol.* 2001;16(12):693–700.
121. Al-Shehbaz IA, O’Kane Jr SL. Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). *Arab B.* 2002;e0001.
122. Pruitt RE, Bowman JL, Grossniklaus U. Plant genetics: a decade of integration. *Nat Genet.* 2003;33:294–304.
123. Koornneef M, Alonso-Blanco C, Vreugdenhil D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu Rev Plant Biol.* 2004;55:141–72.
124. Tonsor SJ, Alonso-Blanco C, Koornneef M. Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in *Arabidopsis thaliana*. *Plant Cell Environ.* 2005;28(1):2–20.
125. Meyerowitz EM. Structure and organization of the *Arabidopsis thaliana* nuclear genome. *COLD SPRING Harb Monogr Ser.* 1994;27:21.
126. Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. *Arabidopsis thaliana*: a model plant for genome analysis. *Science* (80-). 1998;282(5389):662–82.
127. Sharbel TF, Haubold B, Mitchell-Olds T. Genetic isolation by distance in *Arabidopsis thaliana*: Biogeography and postglacial colonization of Europe. *Mol Ecol.* 2000;9(12):2109–18.
128. Hoffmann MH. Biogeography of *Arabidopsis thaliana* (L.) heynh.(Brassicaceae). *J Biogeogr.* 2002;29(1):125–34.

129. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, et al. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 2005;3(7):1289–99.
130. Schmid KJ, Törjék O, Meyer R, Schmuths H, Hoffmann MH, Altmann T. Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theor Appl Genet.* 2006;112(6):1104–14.
131. François O, Blum MGB, Jakobsson M, Rosenberg NA. Demographic history of european populations of *Arabidopsis thaliana*. *PLoS Genet.* 2008;4(5):e1000075.
132. Chory J, Ecker JR, Briggs S, Caboche M, Coruzzi GM, Cook D, et al. National Science Foundation-Sponsored Workshop Report: “The 2010 Project” functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. *Plant Physiol.* 2000;123(2):423–6.
133. Weigel D, Mott R. The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol.* 2009;10(5):107.
134. Kobayashi Y, Weigel D. Move on up, it’s time for change - Mobile signals controlling photoperiod-dependent flowering. *Genes Dev.* 2007;21(19):2371–84.
135. Jones JDG, Dangl JL. The plant immune system. *Nature.* 2006;444(7117):323–9.
136. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat Rev Genet.* 2010;11(8):539–48.
137. Bednarek P. Chemical warfare or modulators of defence responses—the function of secondary metabolites in plant immunity. *Curr Opin Plant Biol.* 2012;15(4):407–14.

138. Maróti G, Kereszt A, Kondorosi E, Mergaert P. Natural roles of antimicrobial peptides in microbes, plants and animals. *Res Microbiol.* 2011;162(4):363–74.
139. Dixon RA. Natural products and plant disease resistance. *Nature.* 2001;411(6839):843–7.
140. Weng J-K, Philippe RN, Noel JP. The rise of chemodiversity in plants. *Science* (80-). 2012;336(6089):1667–70.
141. Hartmann T. The lost origin of chemical ecology in the late 19th century. *Proc Natl Acad Sci.* 2008;105(12):4541–6.
142. Link KP, Angell HR, Walker JC. The isolation of protocatechuic acid from pigmented Onion scales and its relation to disease resistance in Onions. *J Biol Chem.* 1929;81:369–75.
143. D’Auria JC, Gershenzon J. The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Curr Opin Plant Biol.* 2005;8(3):308–16.
144. Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber E, Tokuhsa J, et al. Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. *Phytochemistry.* 2002;59(6):663–71.
145. Chen F, Tholl D, D’Auria JC, Farooq A, Pichersky E, Gershenzon J. Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell.* 2003;15(2):481–94.
146. Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E. Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. *Plant J.* 2005;42(5):757–71.

147. Veit M, Pauli GF. Major flavonoids from *Arabidopsis thaliana* leaves. *J Nat Prod.* 1999;62(9):1301–3.
148. Rohde A, Morreel K, Ralph J, Goeminne G, Hostyn V, De Rycke R, et al. Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell.* 2004;16(10):2749–71.
149. Burchard P, Bilger W, Weissenböck G. Contribution of hydroxycinnamates and flavonoids to epidermal shielding of UV-A and UV-B radiation in developing rye primary leaves as assessed by ultraviolet-induced chlorophyll fluorescence measurements. *Plant Cell Environ.* 2000;23(12):1373–80.
150. Ryan KG, Swinny EE, Winefield C, Markham KR. Flavonoids and UV photoprotection in *Arabidopsis* mutants. *Zeitschrift für Naturforsch C.* 2001;56(9–10):745–54.
151. Fiscus EL, Booker FL. Growth of *Arabidopsis* flavonoid mutant is challenged by radiation longer than the UV-B band. *Environ Exp Bot.* 2002;48(3):213–24.
152. Hagemeyer J, Schneider B, Oldham NJ, Hahlbrock K. Accumulation of soluble and wall-bound indolic metabolites in *Arabidopsis thaliana* leaves infected with virulent or avirulent *Pseudomonas syringae* pathovar tomato strains. *Proc Natl Acad Sci.* 2001;98(2):753–8.
153. Narasimhan K, Basheer C, Bajic VB, Swarup S. Enhancement of plant-microbe interactions using a rhizosphere metabolomics-driven approach and its application in the removal of polychlorinated biphenyls. *Plant Physiol.* 2003;132(1):146–53.

154. Walker TS, Bais HP, Halligan KM, Stermitz FR, Vivanco JM. Metabolic profiling of root exudates of *Arabidopsis thaliana*. *J Agric Food Chem*. 2003;51(9):2548–54.
155. Tan J, Bednarek P, Liu J, Schneider B, Svatoš A, Hahlbrock K. Universally occurring phenylpropanoid and species-specific indolic metabolites in infected and uninfected *Arabidopsis thaliana* roots and leaves. *Phytochemistry*. 2004;65(6):691–9.
156. Van Poecke RMP, Posthumus MA, Dicke M. Herbivore-induced volatile production by *Arabidopsis thaliana* leads to attraction of the parasitoid *Cotesia rubecula*: chemical, behavioral, and gene-expression analysis. *J Chem Ecol*. 2001;27(10):1911–28.
157. Wu C, Dunaway-Mariano D, Mariano PS. Design, Synthesis, and Evaluation of Inhibitors of Pyruvate Phosphate Dikinase. *J Org Chem*. março de 2013;78(5):1910–22.
158. Ward TJ, Clear RM, Rooney AP, O'Donnell K, Gaba D, Patrick S, et al. An adaptive evolutionary shift in *Fusarium* head blight pathogen populations is driving the rapid spread of more toxigenic *Fusarium graminearum* in North America. *Fungal Genet Biol*. 2008;45(4):473–84.
159. Kikot GE, Hours RA, Alconada TM. Review Contribution of cell wall degrading enzymes to pathogenesis of *Fusarium graminearum* : a review. 2009;231–41.
160. Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MK, Brandt ME, et al. One fungus, one name: defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology*. 2013;103(5):400–8.

161. Agrios GN. Plant diseases caused by viruses. *Plant Pathol Fifth Ed Elsevier Acad Press*. 2005;724–820.
162. O’Donnell K, Rooney AP, Proctor RH, Brown DW, McCormick SP, Ward TJ, et al. Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genet Biol*. março de 2013;52:20–31.
163. Nelson PE, Dignani MC, Anaissie EJ. Taxonomy, biology, and clinical aspects of *Fusarium* species. *Clin Microbiol Rev*. 1994;7(4):479–504.
164. O’Donnell K, Kistler HC, Cigelnik E, Ploetz RC. Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene genealogies. *Proc Natl Acad Sci*. 1998;95(5):2044–9.
165. Baayen RP, O’Donnell K, Bonants PJ, Cigelnik E, Kroon LP, Roebroeck EJ, et al. Gene Genealogies and AFLP Analyses in the *Fusarium oxysporum* Complex Identify Monophyletic and Nonmonophyletic *Formae Speciales* Causing Wilt and Rot Disease. *Phytopathology*. 2000;90(8):891–900.
166. Ma L, Geiser DM, Proctor RH, Rooney AP, Donnell KO, Trail F, et al. *Fusarium* Pathogenomics. 2013;
167. Rodriguez-Galvez E, Mendgen K. The infection process of *Fusarium oxysporum* in cotton root tips. *Protoplasma*. 1995;189(1–2):61–72.
168. Wösten HA, Schuren FH, Wessels JG. Interfacial self-assembly of a hydrophobin into an amphipathic protein membrane mediates fungal attachment to

- hydrophobic surfaces. *EMBO J.* 1994;13(24):5848–54.
169. Beswetherick JT, Bishop CD. An ultrastructural study of tomato roots inoculated with pathogenic and non-pathogenic necrotrophic fungi and a saprophytic fungus. *Plant Pathol.* agosto de 1993;42(4):577–88.
170. Lagopodi AL, Ram AFJ, Lamers GEM, Punt PJ, Van den Hondel CAMJJ, Lugtenberg BJJ, et al. Novel aspects of tomato root colonization and infection by *Fusarium oxysporum* f. sp. *radicis-lycopersici* revealed by confocal laser scanning microscopic analysis using the green fluorescent protein as a marker. *Mol Plant Microbe Interact.* 2002;15(2):172–9.
171. Li D, Sirakova T, Rogers L, Ettinger WF, Kolattukudy PE. Regulation of Constitutively Expressed and Induced Cutinase Genes by Different Zinc Finger Transcription Factors in *Fusarium solani* f. sp. *lisi* (*Nectria haematococca*). *J Biol Chem.* 2002;277(10):7905–12.
172. Pérez-Nadales E, Di Pietro A. The membrane mucin Msb2 regulates invasive growth and plant infection in *Fusarium oxysporum*. *Plant Cell.* 2011;23(3):1171–85.
173. Postma J, Rattink H. Biological control of *Fusarium* wilt of carnation with a nonpathogenic isolate of *Fusarium oxysporum*. *Can J Bot.* 1992;70(6):1199–205.
174. Stone JM, Heard JE, Asai T, Ausubel FM. Simulation of fungal-mediated cell death by fumonisin B1 and selection of fumonisin B1-resistant (*fbr*) *Arabidopsis* mutants. *Plant Cell.* 2000;12(10):1811–22.
175. Desmond OJ, Manners JM, Stephens AE, Maclean DJ, Schenk PM, Gardiner

- DM, et al. The Fusarium mycotoxin deoxynivalenol elicits hydrogen peroxide production, programmed cell death and defence responses in wheat. *Mol Plant Pathol.* 2008;9(4):435–45.
176. Jones TM, Anderson AJ, Albersheim P. Host-pathogen interactions IV. Studies on the polysaccharide-degrading enzymes secreted by *Fusarium oxysporum* f. sp. *lycopersici*. *Physiol Plant Pathol.* 1972;2(2):153–66.
177. COOPER RM, WOOD RKS. Induction of Synthesis of Extracellular Cell-wall Degrading Enzymes in Vascular Wilt Fungi. *Nature.* 30 de novembro de 1973;246(5431):309–11.
178. Di Pietro A, Roncero MIG. Endopolygalacturonase from *Fusarium oxysporum* f. sp. *lycopersici*: purification, characterization, and production during infection of tomato plants. *Phytopathology.* 1996;86(12):1324–30.
179. Di Pietro A, Roncero MIG. Purification and characterization of an exopolygalacturonase from the tomato vascular wilt pathogen *Fusarium oxysporum* f. sp. *lycopersici*. *FEMS Microbiol Lett.* 1996;145(2):295–9.
180. García Maceira FI, Di Pietro A, Roncero MIG. Purification and characterization of a novel exopolygalacturonase from *Fusarium oxysporum* f. sp. *lycopersici*. *FEMS Microbiol Lett.* 1997;154(1):37–43.
181. Morrissey JP, Osbourn E. Fungal resistance to plant antibiotics as a mechanism of pathogenesis. *Microbiol Mol Biol Rev.* 1999;63(3):708–24.
182. Berrocal-Lobo M, Molina A. Arabidopsis defense response against *Fusarium oxysporum*. *Trends Plant Sci.* março de 2008;13(3):145–50.

183. Hoffmeister D, Keller NP. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat Prod Rep.* 2007;24(2):393–416.
184. Dufosse L, Fouillaud M, Caro Y, Mapari SAS, Sutthiwong N. Filamentous fungi are large-scale producers of pigments and colorants for the food industry. *Curr Opin Biotechnol.* 2014;26:56–61.
185. Greene GH, McGary KL, Rokas A, Slot JC. Ecology drives the distribution of specialized tyrosine metabolism modules in fungi. *Genome Biol Evol.* 2014;6(1):121–32.
186. Keller NP, Hohn TM. Metabolic Pathway Gene Clusters in Filamentous Fungi. *Fungal Genet Biol.* 1997;21(1):17–29.
187. Hall C, Dietrich FS. The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics.* 2007;177(4):2293–307.
188. Holland PWH. Evolution of homeobox genes. *Wiley Interdiscip Rev Dev Biol.* 2013;2(1):31–45.
189. Wong S, Wolfe KH. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet.* 2005;37(7):777–82.
190. Hittinger CT, Rokas A, Carroll SB. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci U S A.* 2004;101(39):14144–9.
191. Brown DW, McCormick SP, Alexander NJ, Proctor RH, Desjardins AE. A genetic and biochemical approach to study trichothecene diversity in *Fusarium*

- sporotrichioides and *Fusarium graminearum*. *Fungal Genet Biol.* 2001;32(2):121–33.
192. Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature.* 2007;449(7163):677–81.
193. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes. *Science (80-)*. 2012;336(6089):1715–9.
194. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature.* 2004;428(6983):617–24.
195. Ma LJ, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, et al. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* 2009;5(7).
196. Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. *Nature.* 2007;449(7158):54–61.
197. Cornell MJ, Alam I, Soanes DM, Han MW, Hedeler C, Paton NW, et al. Comparative genome analysis across a kingdom of eukaryotic organisms: Specialization and diversification in the Fungi. *Genome Res.* 2007;17(12):1809–22.
198. Hall C, Brachat S, Dietrich FS. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot Cell.* 2005;4(6):1102–15.
199. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ.

- Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell*. 2009;21(7):1897–911.
200. Marcet-Houben M, Gabaldón T. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet*. 2010;26(1):5–8.
 201. League GP, Slot JC, Rokas A. The ASP3 locus in *Saccharomyces cerevisiae* originated by horizontal gene transfer from *Wickerhamomyces*. *FEMS Yeast Res*. 2012;12(7):859–63.
 202. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, et al. Comparative Pathogenomics Reveals Horizontally Acquired Novel Virulence Genes in Fungi Infecting Cereal Hosts. 2012;8(9).
 203. Tiburcio RA, Costa GGL, Carazzolle MF, Mondego JMC, Schuster SC, Carlson JE, et al. Genes acquired by horizontal transfer are potentially involved in the evolution of phytopathogenicity in *moniliophthora perniciosa* and *moniliophthora roreri*, two of the major pathogens of cacao. *J Mol Evol*. 2010;70(1):85–97.
 204. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, et al. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet*. 2006;38(8):953–6.
 205. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. *Curr Biol*. 2006;16(18):1857–64.
 206. Patron NJ, Waller RF, Cozijnsen AJ, Straney DC, Gardiner DM, Nierman WC, et al. Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in

- filamentous ascomycetes. *BMC Evol Biol.* 2007;7:174.
207. Slot JC, Hibbett DS. Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: A phylogenetic study. *PLoS One.* 2007;2(10).
 208. Khaldi N, Collemare J, Lebrun M-H, Wolfe KH. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol.* 2008;9(1):R18.
 209. Slot JC, Rokas A. Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Curr Biol.* 2011;21(2):134–9.
 210. Khaldi N, Wolfe KH. Evolutionary Origins of the Fumonisin Secondary Metabolite Gene Cluster in *Fusarium verticillioides* and *Aspergillus niger*. *Int J Evol Biol.* 2011;2011:423821.
 211. Campbell MA, Rokas A, Slot JC. Horizontal transfer and death of a fungal secondary metabolic gene cluster. *Genome Biol Evol.* 2012;4(3):289–93.
 212. Campbell M a, Staats M, van Kan J a L, Rokas A, Slot JC. Repeated loss of an anciently horizontally transferred gene cluster in *Botrytis*. *Mycologia.* 2013;105(5):1126–34.
 213. Wisecaver JH, Slot JC, Rokas A. The evolution of fungal metabolic pathways. *PLoS Genet.* 2014;10(12):e1004816.
 214. Desjardins AE, Bai GH, Plattner RD, Proctor RH. Analysis of aberrant virulence of *Gibberella zeae* following transformation-mediated complementation of a trichothecene-deficient (Tri5) mutant. *Microbiology.* 2000;146(8):2059–68.

215. Desjardins AE, Hohn TM, McCormick SP. Trichothecene biosynthesis in *Fusarium* species: chemistry, genetics, and significance. *Microbiol Rev.* 1993;57(3):595–604.
216. Brown DW, Butchko RAE, Busman M, Proctor RH. Identification of gene clusters associated with fusaric acid, fusarin, and perithecial pigment production in *Fusarium verticillioides*. *Fungal Genet Biol.* 2012;49(7):521–32.
217. Hansen FT, Sørensen JL, Giese H, Sondergaard TE, Frandsen RJN. Quick guide to polyketide synthase and nonribosomal synthetase genes in *Fusarium*. *Int J Food Microbiol.* 2012;155(3):128–36.
218. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature.* 2010;464(7287):367–73.
219. Bömke C, Tudzynski B. Diversity, regulation, and evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria. *Phytochemistry.* 2009;70(15–16):1876–93.
220. Czymmek KJ, Fogg M, Powell DH, Sweigard J, Park S-Y, Kang S. In vivo time-lapse documentation using confocal and multi-photon microscopy reveals the mechanisms of invasion into the *Arabidopsis* root vascular system by *Fusarium oxysporum*. *Fungal Genet Biol.* 2007;44(10):1011–23.
221. Zhu QH, Stephen S, Kazan K, Jin G, Fan L, Taylor J, et al. Characterization of the defense transcriptome responsive to *Fusarium oxysporum*-infection in *Arabidopsis* using RNA-seq. *Gene.* 2013;512(2):259–66.

222. Lu D, Lin W, Gao X, Wu S, Cheng C, Avila J, et al. Direct ubiquitination of pattern recognition receptor FLS2 attenuates plant innate immunity. *Science*. 2011;332(6036):1439–42.
223. Mudgil Y, Shiu S-H, Stone SL, Salt JN, Goring DR. A large complement of the predicted Arabidopsis ARM repeat proteins are members of the U-box E3 ubiquitin ligase family. *Plant Physiol*. 2004;134(1):59–66.
224. Samuel MA, Salt JN, Shiu S, Goring DR. Multifunctional arm repeat domains in plants. *Int Rev Cytol*. 2006;253:1–26.
225. Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev*. 2002;82(2):373–428.
226. Liu J, Elmore JM, Coaker G. Investigating the functions of the RIN4 protein complex during plant innate immune responses. *Plant Signal Behav*. 2009;4(12):1107–10.
227. Bednarek P, Pislewska-Bednarek M, Svatos A, Schneider B, Doubsky J, Mansurova M, et al. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science*. 2009;323(5910):101–6.
228. Clay NK, Adio AM, Denoux C, Jander G, Ausubel FM. Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science*. 2009;323(5910):95–101.
229. Kidd BN, Kadoo NY, Dombrecht B, Tekeoglu M, Gardiner DM, Thatcher LF, et al. Auxin signaling and transport promote susceptibility to the root-infecting fungal pathogen *Fusarium oxysporum* in Arabidopsis. *Mol Plant Microbe*

- Interact. 2011;24(6):733–48.
230. Pfalz M, Vogel H, Kroymann J. The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*. *Plant Cell*. 2009;21(3):985–99.
 231. Anderson JP, Badruzaufari E, Schenk PM, Manners JM, Desmond OJ, Ehlert C, et al. Antagonistic interaction between abscisic acid and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in *Arabidopsis*. *Plant Cell*. 2004;16(12):3460–79.
 232. Takahashi H, Kanayama Y, Zheng MS, Kusano T, Hase S, Ikegami M, et al. Antagonistic interactions between the SA and JA signaling pathways in *Arabidopsis* modulate expression of defense genes and gene-for-gene resistance to cucumber mosaic virus. *Plant Cell Physiol*. 2004;45(6):803–9.
 233. Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol*. 2005;43:205–27.
 234. Mauch-Mani B, Slusarenko AJ. Systemic acquired resistance in *Arabidopsis thaliana* induced by a predisposing infection with a pathogenic isolate of *Fusarium oxysporum*. *MPMI-Molecular Plant Microbe Interact*. 1994;7(3):378–83.
 235. Edgar CI, McGrath KC, Dombrecht B, Manners JM, Maclean DC, Schenk PM, et al. Salicylic acid mediates resistance to the vascular wilt pathogen *Fusarium oxysporum* in the model host *Arabidopsis thaliana*. *Australas Plant Pathol*. 2006;35(6):581–91.

236. Berrocal-Lobo M, Molina A. Ethylene response factor 1 mediates Arabidopsis resistance to the soilborne fungus *Fusarium oxysporum*. *Mol Plant Microbe Interact.* 2004;17(7):763–70.
237. Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK. Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: Unprecedented sensitivity reveals novel root-and shoot-specific genes. *Plant J.* 2004;38(2):366–79.
238. Kariola T, Brader G, Helenius E, Li J, Heino P, Palva ET. EARLY RESPONSIVE TO DEHYDRATION 15, a negative regulator of abscisic acid responses in Arabidopsis. *Plant Physiol.* 2006;142(4):1559–73.
239. Mauch-Mani B, Mauch F. The role of abscisic acid in plant-pathogen interactions. *Curr Opin Plant Biol.* 2005;8(4):409–14.
240. Melotto M, Underwood W, Koczan J, Nomura K, He SY. Plant Stomata Function in Innate Immunity against Bacterial Invasion. *Cell.* 2006;126(5):969–80.
241. Adie BAT, Pérez-Pérez J, Pérez-Pérez MM, Godoy M, Sánchez-Serrano J-J, Schmelz EA, et al. ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in Arabidopsis. *Plant Cell.* 2007;19(5):1665–81.
242. de Torres-Zabala M, Truman W, Bennett MH, Lafforgue G, Mansfield JW, Egea PR, et al. *Pseudomonas syringae* pv. tomato hijacks the Arabidopsis abscisic acid signalling pathway to cause disease. *EMBO J.* 2007;26(5):1434–43.
243. Lorenzo O, Piqueras R, Sánchez-Serrano JJ, Solano R. ETHYLENE RESPONSE

- FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. *Plant Cell*. 2003;15(1):165–78.
244. Hernández-Blanco C, Feng DX, Hu J, Sánchez-Vallet A, Deslandes L, Llorente F, et al. Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *Plant Cell*. 2007;19(3):890–903.
245. Perfus-Barbeoch L, Jones AM, Assmann SM. Plant heterotrimeric G protein function: Insights from Arabidopsis and rice mutants. *Curr Opin Plant Biol*. 2004;7(6):719–31.
246. Suharsono U, Fujisawa Y, Kawasaki T, Iwasaki Y, Satoh H, Shimamoto K. The heterotrimeric G protein alpha subunit acts upstream of the small GTPase Rac in disease resistance of rice. *Proc Natl Acad Sci U S A*. 2002;99(20):13307–12.
247. Komatsu S, Yang G, Hayashi N, Kaku H, Umemura K, Iwasaki Y. Alterations by a defect in a rice G protein α subunit in probenazole and pathogen-induced responses. *Plant Cell Environ*. 2004;27(7):947–57.
248. Joo JH, Wang S, Chen JG, Jones AM, Fedoroff N V. Different signaling and cell death roles of heterotrimeric G protein alpha and beta subunits in the Arabidopsis oxidative stress response to ozone. *Plant Cell*. 2005;17(3):957–70.
249. Olsen AN, Ernst HA, Leggio LL, Skriver K. NAC transcription factors: Structurally distinct, functionally diverse. *Trends Plant Sci*. 2005;10(2):79–87.
250. Delessert C, Kazan K, Wilson IW, Straeten D Van Der, Manners J, Dennis ES, et al. The transcription factor ATAF2 represses the expression of pathogenesis-related genes in Arabidopsis. *Plant J*. 2005;43(5):745–57.

251. Nakano T, Suzuki K, Ohtsuki N, Tsujimoto Y, Fujimura T, Shinshi H. Identification of genes of the plant-specific transcription-factor families cooperatively regulated by ethylene and jasmonate in *Arabidopsis thaliana*. *J Plant Res.* 2006;119(4):407–13.
252. Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol.* 2006;140(2):411–32.
253. McGrath KC, Dombrecht B, Manners JM, Schenk PM, Edgar CI, Maclean DJ, et al. Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiol.* 2005;139(2):949–59.
254. Epple P, Apel K, Bohlmann H. Overexpression of an endogenous thionin enhances resistance of *Arabidopsis* against *Fusarium oxysporum*. *Plant Cell.* 1997;9(4):509–20.
255. Chan YL, Prasad V, Sanjaya, Kuei HC, Po CL, Chan MT, et al. Transgenic tomato plants expressing an *Arabidopsis* thionin (Thi2.1) driven by fruit-inactive promoter battle against phytopathogenic attack. *Planta.* 2005;221(3):386–93.
256. Takemoto D, Tanaka A, Scott B. A p67Phox-like regulator is recruited to control hyphal branching in a fungal-grass mutualistic symbiosis. *Plant Cell.* 2006;18(10):2807–21.
257. Schwachtje J, Minchin PEH, Jahnke S, van Dongen JT, Schittko U, Baldwin IT. SNF1-related kinases allow plants to tolerate herbivory by allocating carbon to roots. *Proc Natl Acad Sci U S A.* 2006;103(34):12935–40.

258. Llorente F, Alonso-Blanco C, Sánchez-Rodríguez C, Jorda L, Molina A. ERECTA receptor-like kinase and heterotrimeric G protein from Arabidopsis are required for resistance to the necrotrophic fungus *Plectosphaerella cucumerina*. *Plant J.* 2005;43(2):165–80.
259. Day RB, Tanabe S, Koshioka M, Mitsui T, Itoh H, Ueguchi-Tanaka M, et al. Two rice GRAS family genes responsive to N-acetylchitoooligosaccharide elicitor are induced by phytoactive gibberellins: Evidence for cross-talk between elicitor and gibberellin signaling in rice cells. *Plant Mol Biol.* 2004;54(2):261–72.
260. Izaguirre MM, Mazza CA, Biondini M, Baldwin IT, Ballaré CL. Remote sensing of future competitors: impacts on plant defenses. *Proc Natl Acad Sci U S A.* 2006;103(18):7170–4.
261. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol.* 2010;28(3):245–8.
262. Joyce AR, Palsson BØ. Predicting gene essentiality using genome-scale in silico models. *Microb Gene Essentiality Protoc Bioinforma.* 2008;433–57.
263. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, et al. Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Mol Syst Biol.* 2011;7(1):460.
264. Edwards JS, Covert M, Palsson B. Metabolic modelling of microbes: the flux-balance approach. *Environ Microbiol.* 2002;4(3):133–40.
265. Reed JL, Palsson BØ. Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that

- comprise network states. *Genome Res.* 2004;14(9):1797–805.
266. Cuevas DA, Edirisinghe J, Henry CS, Overbeek R, O’Connell TG, Edwards RA. From DNA to FBA: how to build your own genome-scale metabolic model. *Front Microbiol.* 2016;7.
267. Price ND, Reed JL, Palsson BØ. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol.* 2004;2(11):886–97.
268. Feist AM, Palsson BO. The biomass objective function. *Curr Opin Microbiol.* 2010;13(3):344–9.
269. Varma A, Boesch BW, Palsson BO. Biochemical production capabilities of *Escherichia coli*. *Biotechnol Bioeng.* 1993;42(1):59–73.
270. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 2003;13(2):244–53.
271. Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am J Physiol Integr Comp Physiol.* 2001;280(3):R695–704.
272. Peyraud R, Dubiella U, Barbacci A, Genin S, Raffaele S, Roby D. Advances on plant–pathogen interactions from molecular toward systems biology perspectives. *Plant J.* 2017;90(4):720–37.
273. Catharina, L; De Menezes, MA; Carels N. System biology to access target relevance in the research and development of molecular inhibitors. In:

Theoretical and Applied Aspects of Systems Biology. 2018.

274. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. *Nucleic Acids Res.* 2014, 42. D749–55 [CrossRef],[PubMed],[CAS].
275. Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: A functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 2009;38(SUPPL.1).
276. The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2014;42:D191--8.
277. Sigrist CJA, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res.* 2012;41(D1):D344–7.
278. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2013;42(D1):D222–30.
279. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, et al. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 2006;35(suppl_1):D237–40.
280. Simossis VA, Heringa J. PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* 2005;33(SUPPL. 2).
281. Rost B. Twilight zone of protein sequence alignments. *Protein Eng.* 1999;12(2):85–94.
282. Catharina L, Carlyle Ribeiro Lima AF, Guimarães ACR, Alves-Ferreira M,

- Tuffery P, Derreumaux P, et al. A Computational Methodology to Overcome the Challenges Associated With the Search for Specific Enzyme Targets to Develop Drugs Against. *Bioinform Biol Insights*. 2017;11:1–21.
283. Néron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, et al. Mobylye: a new full web bioinformatics framework. *Bioinformatics*. 2009;25(22):3005–11.
284. Söding J. Protein homology detection by HMM–HMM comparison. *Bioinformatics*. 2004;21(7):951–60.
285. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 1 de julho de 2005;33:W244-8.
286. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 25 de dezembro de 2011;9(2):173–5.
287. Lima CR, Carels N, Guimaraes ACR, Tufféry P, Derreumaux P. In silico structural characterization of protein targets for drug development against *Trypanosoma cruzi*. *J Mol Model*. 2016;22(10):244.
288. Panigrahi AK, Schnauffer A, Ernst NL, Wang B, Carmean N, Salavati R, et al. Identification of novel components of *Trypanosoma brucei* editosomes. *Rna*. 2003;9(4):484–92.
289. Panigrahi AK, Gygi SP, Ernst NL, Igo RP, Palazzo SS, Schnauffer A, et al. Association of two novel proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA editing complex. *Mol Cell Biol*. 2001;21(2):380–9.

290. Mehta V, Sen R, Moshiri H, Salavati R. Mutational analysis of *Trypanosoma brucei* RNA editing ligase reveals regions critical for interaction with KREPA2. *PLoS One*. 2015;10(3):e0120844.
291. Simpson AGB, Stevens JR, Lukeš J. The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol*. 2006;22(4):168–74.
292. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, et al. OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res*. 2016;45(D1):D744–9.
293. Pitkänen E, Jouhten P, Hou J, Syed MF, Blomberg P, Kludas J, et al. Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. Nielsen J, organizador. *PLoS Comput Biol*. 6 de fevereiro de 2014;10(2):e1003465.
294. Yu J, Picord G, Tuffery P, Guerois R. HAlign-Kbest: exploring sub-optimal alignments for remote homology comparative modeling. *Bioinformatics*. 2015;31(23):3850–2.
295. Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. Protein structure modeling with MODELLER. *Struct proteomics high-throughput methods*. 2008;145–59.
296. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004;32(suppl_2):W665–7.
297. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of

- a protein. *J Mol Biol.* 1982;157(1):105–32.
298. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605–12.
299. Heger A, Mallick S, Wilton C, Holm L. The global trace graph, a novel paradigm for searching protein sequence databases. *Bioinformatics.* 2007;23(18):2361–7.
300. de Oliveira Dal’Molin CG, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK. AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. *Plant Physiol.* fevereiro de 2010;152(2):579–89.
301. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26(2):283–91.
302. Eisenberg D, Lüthy R, Bowie JU. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* 1997;277:396–404.
303. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35(suppl_2):W407–10.
304. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr.* 2010;66(1):12–21.
305. Wilming LG, Gilbert JGR, Howe K, Trevanion S, Hubbard T, Harrow JL. The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.*

- 2007;36(suppl_1):D753–60.
306. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2012. *Nucleic Acids Res.* 2011;40(D1):D84–90.
 307. Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, Hahn MW. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput Biol.* 2014;10(12):e1003998.
 308. Phillippy AM, Schatz MC, Pop M. Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol.* 2008;9(3):R55.
 309. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. *Nat Methods.* 2011;8(1):61–5.
 310. Han M V, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* 2013;30(8):1987–97.
 311. Echols N, Harrison P, Balasubramanian S, Luscombe NM, Bertone P, Zhang Z, et al. Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Res.* 2002;30(11):2515–23.
 312. Balakirev ES, Ayala FJ. Pseudogenes: are they “junk” or functional DNA? *Annu Rev Genet.* 2003;37:123–51.
 313. Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu X, et al. The GENCODE pseudogene resource. *Genome Biol.* 2012;13(9):R51.

314. Selkov E, Grechkin Y, Mikhailova N, Selkov E. MPW: The Metabolic Pathways Database. *Nucleic Acids Res.* 1998;26(1):43–5.
315. de Souza W, Rodrigues JCF. Sterol Biosynthesis Pathway as Target for Anti-trypanosomatid Drugs. *Interdiscip Perspect Infect Dis.* 2009;1–19.
316. Xu W, Hsu F-F, Baykal E, Huang J, Zhang K. Sterol Biosynthesis Is Required for Heat Resistance but Not Extracellular Survival in *Leishmania*. *PLoS Pathog.* 2014;10:e1004427.
317. McCall L-I, El Aroussi A, Choi JY, Vieira DF, De Muylder G, Johnston JB, et al. Targeting Ergosterol Biosynthesis in *Leishmania donovani*: Essentiality of Sterol 14 α -demethylase. *PLoS Negl Trop Dis.* 2015;9:e0003588.
318. Roberts CW, McLeod R, Rice DW, Ginger M, Chance ML, Goad LJ. Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Mol Biochem Parasitol.* fevereiro de 2003;126(2):129–42.
319. Fairlamb AH, Blackburn P, Ulrich P, Chait BT, Cerami A. Trypanothione: a novel bis (glutathionyl) spermidine cofactor for glutathione reductase in trypanosomatids. *Science (80-).* 1985;227(4693):1485–7.
320. Leroux AE, Krauth-Siegel RL. Thiol redox biology of trypanosomatids and potential targets for chemotherapy. *Mol Biochem Parasitol.* março de 2016;206(1–2):67–74.
321. Bringaud F, Baltz D, Baltz T. Functional and molecular characterization of a glycosomal PPi-dependent enzyme in trypanosomatids: pyruvate, phosphate

- dikinase. *Proc Natl Acad Sci*. 1998;95(14):7963–8.
322. Deramchia K, Morand P, Biran M, Millerioux Y, Mazet M, Wagnies M, et al. Contribution of Pyruvate Phosphate Dikinase in the Maintenance of the Glycosomal ATP/ADP Balance in the *Trypanosoma brucei* Procyclic Form. *J Biol Chem*. 20 de junho de 2014;289(25):17365–78.
323. Rodriguez-Contreras D, Aslan H, Feng X, Tran K, Yates PA, Kamhawi S, et al. Regulation and biological function of a flagellar glucose transporter in *Leishmania mexicana*: a potential glucose sensor. *FASEB J*. 2015;29(1):11–24.
324. Gao G, Simpson AM, Kang X, Rogers K, Nebohacova M, Li F, et al. Functional complementation of *Trypanosoma brucei* RNA in vitro editing with recombinant RNA ligase. *Proc Natl Acad Sci U S A*. 2005;102(13):4712–7.
325. Kang X, Gao G, Rogers K, Falick AM, Zhou S, Simpson L. Reconstitution of full-round uridine-deletion RNA editing with three recombinant proteins. *Proc Natl Acad Sci*. 2006;103(38):13944–9.
326. Schnauffer A, Panigrahi AK, Panicucci B, Igo RP, Salavati R, Stuart K. An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science* (80-). 2001;291(5511):2159–62.
327. Gao G, Simpson L. Is the *Trypanosoma brucei* REL1 RNA ligase specific for U-deletion RNA editing, and is the REL2 RNA ligase specific for U-insertion editing? *J Biol Chem*. 2003;278(30):27570–4.
328. Crowther GJ, Shanmugam D, Carmona SJ, Doyle MA, Hertz-Fowler C, Berriman M, et al. Identification of attractive drug targets in neglected-disease

- pathogens using an in silico approach. *PLoS Negl Trop Dis*. 2010;4(8):e804.
329. Flórez AF, Park D, Bhak J, Kim B-C, Kuchinsky A, Morris JH, et al. Protein network prediction and topological analysis in *Leishmania major* as a tool for drug target selection. *BMC Bioinformatics*. 2010;11(1):484.
330. Nes WD, Zhou W, Ganapathy K, Liu J, Vatsyayan R, Chamala S, et al. Sterol 24-C-methyltransferase: an enzymatic target for the disruption of ergosterol biosynthesis and homeostasis in *Cryptococcus neoformans*. *Arch Biochem Biophys*. 2009;481(2):210–8.
331. Goto Y, Bhatia A, Raman VS, Vidal SEZ, Bertholet S, Coler RN, et al. *Leishmania infantum* sterol 24-c-methyltransferase formulated with MPL-SE induces cross-protection against *L. major* infection. *Vaccine*. 2009;27(21):2884–90.
332. Goto Y, Bogatzki LY, Bertholet S, Coler RN, Reed SG. Protective immunization against visceral leishmaniasis using *Leishmania* sterol 24-c-methyltransferase formulated in adjuvant. *Vaccine*. 2007;25(42):7450–8.
333. Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov*. 2008;7(11):900–7.
334. Chacko AD, McDade SS, Chanduloy S, Church SW, Kennedy R, Price J, et al. Expression of the SEPT9_i4 isoform confers resistance to microtubule-interacting drugs. *Cell Oncol*. 26 de abril de 2012;35(2):85–93.
335. Bury M, Andolfi A, Rogister B, Cimmino A, Mégalizzi V, Mathieu V, et al.

- Fusicoccin a, a phytotoxic carbocyclic diterpene glucoside of fungal origin, reduces proliferation and invasion of glioblastoma cells by targeting multiple tyrosine kinases. *Transl Oncol.* abril de 2013;6(2):112–23.
336. Rispaill N, Dita M, González-Verdejo C, Pérez-de-Luque A, Castillejo M, Prats E, et al. Plant resistance to parasitic plants: molecular approaches to an old foe. *New Phytol.* 2007;173(4):703–12.
337. Keane PJ. Horizontal or Generalized Resistance to Pathogens in Plants. In: *Plant Pathology.* InTech; 2012.
338. Flor HH. Current status of the gene-for-gene concept. *Annu Rev Phytopathol.* 1971;9(1):275–96.
339. Arteaga CL, Engelman JA. ERBB Receptors: From Oncogene Discovery to Basic Science to Mechanism-Based Cancer Therapeutics. *Cancer Cell.* março de 2014;25(3):282–303.
340. Magaraci F, Jimenez, Rodrigues C, Rodrigues JCF, Braga MV, Yardley V, et al. Azasterols as Inhibitors of Sterol 24-Methyltransferase in *Leishmania* Species and *Trypanosoma cruzi*. *J Med Chem.* outubro de 2003;46(22):4714–27.
341. Amaro RE, Schnauffer A, Interthal H, Hol W, Stuart KD, McCammon JA. Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proc Natl Acad Sci.* 2008;105(45):17278–83.
342. Moshiri H, Acoca S, Kala S, Najafabadi HS, Hogues H, Purisima E, et al. Naphthalene-based RNA editing inhibitor blocks RNA editing activities and editosome assembly in *Trypanosoma brucei*. *J Biol Chem.* 2011;286(16):14178–

89.

343. Liang S, Connell GJ. Identification of specific inhibitors for a trypanosomatid RNA editing reaction. *RNA*. 2010;16(12):2435–41.
344. Herzberg O, Chen CCH, Liu S, Tempczyk A, Howard A, Wei M, et al. Pyruvate site of pyruvate phosphate dikinase: crystal structure of the enzyme-phosphonopyruvate complex, and mutant analysis. *Biochemistry*. 22 de janeiro de 2002;41(3):780–7.
345. Palayam M, Lakshminarayanan K, Radhakrishnan M, Krishnaswamy G. Preliminary analysis to target pyruvate phosphate dikinase from wolbachia endosymbiont of *Brugia malayi* for designing anti-filarial agents. *Interdiscip Sci Comput Life Sci*. 6 de março de 2012;4(1):74–82.
346. Benítez D, Medeiros A, Fiestas L, Panozzo-Zenere EA, Maiwald F, Prousis KC, et al. Identification of Novel Chemical Scaffolds Inhibiting Trypanothione Synthetase from Pathogenic Trypanosomatids. *PLoS Negl Trop Dis*. 12 de abril de 2016;10(4):e0004617.
347. Kim KM, Caetano-Anollés G. The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol Biol*. 2011;11(1):140.
348. Oza SL, Shaw MP, Wyllie S, Fairlamb AH. Trypanothione biosynthesis in *Leishmania major*. *Mol Biochem Parasitol*. 2005;139(1):107–16.
349. Leifso K, Cohen-Freue G, Dogra N, Murray A, McMaster WR. Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: the *Leishmania* genome is constitutively expressed. *Mol Biochem*

- Parasitol. 2007;152(1):35–46.
350. Opperdoes FR, Coombs GH. Metabolism of *Leishmania*: proven and predicted. *Trends Parasitol.* 2007;23(4):149–58.
351. Walkowiak S, Rowland O, Rodrigue N, Subramaniam R. Whole genome sequencing and comparative genomics of closely related *Fusarium* Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genomics.* 2016;17(1):1014.
352. Bukhalid RA, Takeuchi T, Labeda D, Loria R. Horizontal transfer of the plant virulence gene, *necl*, and flanking sequences among genetically distinct *Streptomyces* strains in the *Diastatochromogenes* cluster. *Appl Environ Microbiol.* 2002;68(2):738–44.
353. Burger G, Lang BF. Parallels in genome evolution in mitochondria and bacterial symbionts. *IUBMB Life.* 2003;55(4-5):205–12.
354. Loreto ELS, Carareto CMA, Capy P. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity (Edinb).* 2008;100(6):545–54.
355. Trobridge GD. Foamy virus vectors for gene transfer. *Expert Opin Biol Ther.* 2009;9(11):1427–36.
356. Haegeman A, Jones JT, Danchin EGJ. Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol Plant-Microbe Interact.* 2011;24(8):879–87.
357. Gao C, Ren X, Mason AS, Liu H, Xiao M, Li J, et al. Horizontal gene transfer in plants. *Funct Integr Genomics.* 2014;14(1):23–9.

358. Emiliani G, Fondi M, Fani R, Gribaldo S. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biol Direct*. 2009;4(1):7.
359. Tzin V, Galili G. The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *Arab B*. 2010;e0132.
360. Cabib E, Roberts R, Bowers B. Synthesis of the yeast cell wall and its regulation. *Annu Rev Biochem*. 1982;51(1):763–93.
361. Behr J. Chitin synthase as an antifungal target: recent advances. *Curr Med Chem Agents*. 2003;2(2):173–89.
362. Ruiz-Herrera J, San-Blas G. Chitin synthesis as a target for antifungal drugs. *Curr Drug Targets-Infectious Disord*. 2003;3(1):77–91.
363. Jiang H, Wang S, Dang L, Wang S, Chen H, Wu Y. A Novel Short-Root Gene Encodes a Glucosamine-6-Phosphate Acetyltransferase Required for Maintaining Normal Root Cell Shape in Rice 1. 2005;138(May):232–42.
364. Nozaki M, Sugiyama M, Duan J, Uematsu H, Genda T, Sato Y. A Missense Mutation in the Glucosamine-6-Phosphate N -Acetyltransferase – Encoding Gene Causes Temperature-Dependent Growth Defects and Ectopic Lignin Deposition in *Arabidopsis*. 2012;24(August):3366–79.
365. Chen Y-H, Shen H-L, Hsu P-J, Hwang S-G, Cheng W-H. N-Acetylglucosamine-1-P uridylyltransferase 1 and 2 are required for gametogenesis and embryo development in *Arabidopsis thaliana*. *Plant Cell Physiol*. 2014;55(11):1977–93.
366. M Chaudhary P, G Tupe S, V Deshpande M. Chitin synthase inhibitors as

- antifungal agents. *Mini Rev Med Chem*. 2013;13(2):222–36.
367. Graybill JR, Najvar LK, Bocanegra R, Hector RF, Luther MF. Efficacy of nikkomycin Z in the treatment of murine histoplasmosis. *Antimicrob Agents Chemother*. 1998;42(9):2371–4.
368. Li RK, Rinaldi MG. In vitro antifungal activity of nikkomycin Z in combination with fluconazole or itraconazole. *Antimicrob Agents Chemother*. 1999;43(6):1401–5.
369. Stevens DA. Drug interaction studies of a glucan synthase inhibitor (LY 303366) and a chitin synthase inhibitor (Nikkomycin Z) for inhibition and killing of fungal pathogens. *Antimicrob Agents Chemother*. 2000;44(9):2547–8.
370. Behr J-B, Gautier-Lefebvre I, Mvondo-Evina C, Guillerm G, Ryder NS. Inhibition of chitin synthetase from *Saccharomyces cerevisiae* by a new UDP-GlcNAc analogue. *J Enzyme Inhib*. 2001;16(2):107–12.
371. Winn M, Goss RJM, Kimura K, Bugg TDH. Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure–function studies and nucleoside biosynthesis. *Nat Prod Rep*. 2010;27(2):279–304.
372. Ge Z, Ji Q, Chen C, Liao Q, Wu H, Liu X, et al. Synthesis and biological evaluation of novel 3-substituted amino-4-hydroxylcoumarin derivatives as chitin synthase inhibitors and antifungal agents. *J Enzyme Inhib Med Chem*. 3 de março de 2016;31(2):219–28.
373. Wang Z, Kiledjian M. Functional link between the mammalian exosome and mRNA decapping. *Cell*. 2001;107(6):751–62.

374. Liu H, Rodgers ND, Jiao X, Kiledjian M. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *EMBO J.* 2002;21(17):4699–708.
375. Jarecki J, Chen X, Bernardino A, Coovert DD, Whitney M, Burghes A, et al. Diverse small-molecule modulators of SMN expression found by high-throughput compound screening: early leads towards a therapeutic for spinal muscular atrophy. *Hum Mol Genet.* 2005;14(14):2003–18.
376. Singh J, Salcius M, Liu S-W, Staker BL, Mishra R, Thurmond J, et al. DcpS as a Therapeutic Target for Spinal Muscular Atrophy. *ACS Chem Biol.* 21 de novembro de 2008;3(11):711–22.
377. Gopalsamy A, Narayanan A, Liu S, Parikh MD, Kyne Jr RE, Fadeyi O, et al. Design of Potent mRNA Decapping Scavenger Enzyme (DcpS) Inhibitors with Improved Physicochemical Properties To Investigate the Mechanism of Therapeutic Benefit in Spinal Muscular Atrophy (SMA). *J Med Chem.* 2017;60(7):3094–108.
378. Zhou H, Lorenz MC. Carnitine acetyltransferases are required for growth on non-fermentable carbon sources but not for pathogenesis in *Candida albicans*. *Microbiology.* 2008;154(2):500–9.
379. Strijbis K, van Roermund CW, van den Burg J, van den Berg M, Hardy GPM, Wanders RJ, et al. Contributions of Carnitine Acetyltransferases to Intracellular Acetyl Unit Transport in *Candida albicans*. *J Biol Chem.* 6 de agosto de 2010;285(32):24335–46.
380. Kunau W-H, Dommes V, Schulz H. β -Oxidation of fatty acids in mitochondria,

- peroxisomes, and bacteria: a century of continued progress. *Prog Lipid Res.* 1995;34(4):267–342.
381. Van Roermund CW, Elgersma Y, Singh N, Wanders RJ, Tabak HF. The membrane of peroxisomes in *Saccharomyces cerevisiae* is impermeable to NAD (H) and acetyl-CoA under in vivo conditions. *EMBO J.* 1995;14(14):3480.
382. Bhambra GK, Wang Z, Soanes DM, Wakley GE, Talbot NJ. Peroxisomal carnitine acetyl transferase is required for elaboration of penetration hyphae during plant infection by *Magnaporthe grisea*. *Mol Microbiol.* 2006;61(1):46–60.
383. Franken J, Kroppenstedt S, Swiegers JH, Bauer FF. Carnitine and carnitine acetyltransferases in the yeast *Saccharomyces cerevisiae*: a role for carnitine in stress protection. *Curr Genet.* 22 de junho de 2008;53(6):347–60.
384. Son H, Min K, Lee J, Choi GJ, Kim J-C, Lee Y-W. Mitochondrial carnitine-dependent acetyl coenzyme A transport is required for normal sexual and asexual development of the ascomycete *Gibberella zeae*. *Eukaryot Cell.* 2012;11(9):1143–53.
385. Jaudzems K, Kuka J, Gutsaits A, Zinovjevs K, Kalvinsh I, Liepinsh E, et al. Inhibition of carnitine acetyltransferase by mildronate, a regulator of energy metabolism. *J Enzyme Inhib Med Chem.* 16 de dezembro de 2009;24(6):1269–75.
386. Zhang Y, Colabroy KL, Begley TP, Ealick SE. Structural studies on 3-hydroxyanthranilate-3, 4-dioxygenase: the catalytic mechanism of a complex oxidation involved in NAD biosynthesis. *Biochemistry.* 2005;44(21):7632–43.

387. Stone TW, Darlington LG. Endogenous kynurenines as targets for drug discovery and development. *Nat Rev Drug Discov.* 2002;1(8):609–20.
388. McNinch JK, McNinch JD, May SW. Catalysis, stereochemistry, and inhibition of ureidoglycolate lyase. *J Biol Chem.* 2003;278(50):50091–100.
389. Dumont ME, Cardillo TS, Hayes MK, Sherman F. Role of cytochrome c heme lyase in mitochondrial import and accumulation of cytochrome c in *Saccharomyces cerevisiae*. *Mol Cell Biol.* novembro de 1991;11(11):5487–96.
390. Mayer A, Neupert W, Lill R. Translocation of apocytochrome c across the outer membrane of mitochondria. *J Biol Chem.* 1995;270(21):12390–7.
391. Nicholson DW, Köhler H, Neupert W. Import of cytochrome c into mitochondria. *FEBS J.* 1987;164(1):147–57.
392. Hofmann K, Tomiuk S, Wolff G, Stoffel W. Cloning and characterization of the mammalian brain-specific, Mg²⁺-dependent neutral sphingomyelinase. *Proc Natl Acad Sci.* 2000;97(11):5895–900.
393. Nagiec MM, Nagiec EE, Baltisberger JA, Wells GB, Lester RL, Dickson RC. Sphingolipid synthesis as a target for antifungal drugs complementation of the inositol phosphorylceramide synthase defect in a mutant strain of *Saccharomyces cerevisiae* by the AUR1 gene. *J Biol Chem.* 1997;272(15):9809–17.
394. Mandala SM, Thornton RA, Frommer BR, Curotto JE, Rozdilsky W, Kurtz MB, et al. The discovery of australifungin, a novel inhibitor of sphinganine N-acyltransferase from *Sporormiella australis*. *J Antibiot (Tokyo).* 1995;48(5):349–56.

395. Domin N, Wilson D, Brock M. Methylcitrate cycle activation during adaptation of *Fusarium solani* and *Fusarium verticillioides* to propionyl-CoA-generating carbon sources. *Microbiology*. dezembro de 2009;155(Pt 12):3903–12.
396. Völkl A, Vogler B, Schollenberger M, Karlovsky P. Microbial detoxification of mycotoxin deoxynivalenol. *J Basic Microbiol*. 2004;44(2):147–56.
397. De Lucca AJ. Harmful fungi in both agriculture and medicine. *Rev Iberoam Micol*. março de 2007;24(1):3–13.
398. Jestoi M. Emerging *Fusarium*-mycotoxins fusaproliferin, beauvericin, enniatins, and moniliformin—A review. *Crit Rev Food Sci Nutr*. 2008;48(1):21–49.
399. Brock M, Buckel W. On the mechanism of action of the antifungal agent propionate. *Eur J Biochem*. agosto de 2004;271(15):3227–41.
400. Yuasa HJ, Ball HJ. Efficient tryptophan-catabolizing activity is consistently conserved through evolution of TDO enzymes, but not IDO enzymes. *J Exp Zool Part B Mol Dev Evol*. 2015;324(2):128–40.
401. Yuasa HJ, Ball HJ. Indoleamine 2, 3-dioxygenases with very low catalytic activity are well conserved across kingdoms: IDOs of Basidiomycota. *Fungal Genet Biol*. 2013;56:98–106.
402. Moon YW, Hajjar J, Hwu P, Naing A. Targeting the indoleamine 2, 3-dioxygenase pathway in cancer. *J Immunother cancer*. 2015;3(1):51.
403. Lovelace MD, Varney B, Sundaram G, Franco NF, Ng ML, Pai S, et al. Current evidence for a role of the kynurenine pathway of tryptophan metabolism in multiple sclerosis. *Front Immunol*. 2016;7.

404. R Selvan S, P Dowling J, K Kelly W, Lin J. Indoleamine 2, 3-dioxygenase (IDO): biology and target in cancer immunotherapies. *Curr Cancer Drug Targets*. 2016;16(9):755–64.
405. Brochez L, Chevolet I, Kruse V. The rationale of indoleamine 2, 3-dioxygenase inhibition for cancer therapy. *Eur J Cancer*. 2017;76:167–82.
406. Song P, Ramprasath T, Wang H, Zou M-H. Abnormal kynurenine pathway of tryptophan catabolism in cardiovascular diseases. *Cell Mol Life Sci*. 2017;1–18.
407. Singh T, Goel RK. Adjuvant indoleamine 2, 3-dioxygenase enzyme inhibition for comprehensive management of epilepsy and comorbid depression. *Eur J Pharmacol*. 2016;784:111–20.
408. Markwalder JA, Seitz SP, Blat Y, Elkin L, Hunt JT, Pabalan JG, et al. Identification and optimization of a novel series of indoleamine 2, 3-dioxygenase inhibitors. *Bioorg Med Chem Lett*. 2017;27(3):582–5.
409. Jang J-P, Jang J-H, Soung N-K, Kim H-M, Jeong S-J, Asami Y, et al. Benzomalvin E, an indoleamine 2, 3-dioxygenase inhibitor isolated from *Penicillium* sp. FN070315. *J Antibiot (Tokyo)*. 2012;65(4):215.
410. Jang J-P, Jang J-H, Oh M, Son S, Kim SM, Kim H-M, et al. Inhibition of indoleamine 2, 3-dioxygenase by thielavin derivatives from a soil fungus, *Coniochaeta* sp. 10F058. *J Antibiot (Tokyo)*. 2014;67(4):331.
411. Kodaki T, Yamashita S. Yeast phosphatidylethanolamine methylation pathway. Cloning and characterization of two distinct methyltransferase genes. *J Biol Chem*. 1987;262(32):15428–35.

412. McGraw P, Henry SA. Mutations in the *Saccharomyces cerevisiae* *opi3* gene: effects on phospholipid methylation, growth and cross-pathway regulation of inositol synthesis. *Genetics*. 1989;122(2):317–30.
413. Markham P, Bainbridge BW. Characterization of a new choline locus in *Aspergillus nidulans* and its significance for choline metabolism. *Genet Res (Camb)*. 1978;32(3):303–10.
414. Robson GD, Wiebe MG, Cunliffe B, Trinci APJ. Choline-and acetylcholine-induced changes in the morphology of *Fusarium graminearum*: evidence for the involvement of the choline transport system and acetylcholinesterase. *Microbiology*. 1995;141(6):1309–14.
415. Plateau P, Fromant M, Schmitter JM, Blanquet S. Catabolism of bis (5'-nucleosidyl) tetraphosphates in *Saccharomyces cerevisiae*. *J Bacteriol*. 1990;172(12):6892–9.
416. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*. 2009;7(2):129–43.
417. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*. 2009;5(1):320.
418. Kumar VS, Dasika MS, Maranas CD. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*. 2007;8(1):212.
419. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*. 2012;10(4):291–305.

420. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet.* 2014;15(2):107–20.
421. O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. *Cell.* 2015;161(5):971–87.
422. Boyle NR, Morgan JA. Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. *BMC Syst Biol.* 2009;3(1):4.
423. Feist AM, Palsson BØ. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol.* 2008;26(6):659–67.
424. McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol Syst Biol.* 2013;9(1):661.
425. Baart GJE, Zomer B, de Haan A, van der Pol LA, Beuvery EC, Tramper J, et al. Modeling *Neisseria meningitidis* metabolism: from genome to metabolic fluxes. *Genome Biol.* 2007;8(7):R136.
426. Duarte NC, Herrgård MJ, Palsson BØ. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 2004;14(7):1298–309.
427. Becker SA, Palsson BØ. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.* 2005;5(1):8.
428. Pastink MI, Teusink B, Hols P, Visser S, de Vos WM, Hugenholtz J. Genome-

- scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria. *Appl Environ Microbiol.* 2009;75(11):3627–33.
429. Pramanik J, Keasling JD. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol Bioeng.* 1997;56(4):398–421.
430. Schwender J, Hay JO. Predictive modeling of biomass component tradeoffs in *Brassica napus* developing oilseeds based on in silico manipulation of storage metabolism. *Plant Physiol.* 2012;pp-112.
431. Novak L, Loubiere P. The metabolic network of *Lactococcus lactis*: distribution of ¹⁴C-labeled substrates between catabolic and anabolic pathways. *J Bacteriol.* 2000;182(4):1136–43.
432. Hay J, Schwender J. Metabolic network reconstruction and flux variability analysis of storage synthesis in developing oilseed rape (*Brassica napus* L.) embryos. *Plant J.* 2011;67(3):526–41.
433. Collakova E, Yen JY, Senger RS. Are we ready for genome-scale modeling in plants? *Plant Sci.* 2012;191:53–70.
434. Yuan H, Cheung CY, Hilbers PAJ, van Riel NAW. Flux balance analysis of plant metabolism: the effect of biomass composition and model structure on model predictions. *Front Plant Sci.* 2016;7:537.
435. Poolman MG, Miguet L, Sweetlove LJ, Fell D a. A genome-scale metabolic model of *Arabidopsis* and some of its properties. *Plant Physiol.* novembro de 2009;151(3):1570–81.

436. Arnold A, Nikoloski Z. Bottom-up metabolic reconstruction of *Arabidopsis thaliana* and its application to determining the metabolic costs of enzyme production. *Plant Physiol.* 2014;pp-114.
437. Ohlrogge J, Browse J. Lipid biosynthesis. *Plant Cell.* 1995;7(7):957.
438. Mueller LA, Zhang P, Rhee SY. AraCyc : A Biochemical Pathway Database for *Arabidopsis* 1. 2003;132(June):453–60.
439. Moore D. David Moore's World of Fungi: Where Mycology Starts. *Ericoid Mycorrhizas.* 2013;
440. Christias C, Couvaraki C, Georgopoulos SG, Macris B, Vomvoyanni V. Protein content and amino acid composition of certain fungi evaluated for microbial protein production. *Appl Microbiol.* 1975;29(2):250–4.
441. Srivastava S, Pathak N, Srivastava P. Identification of limiting factors for the optimum growth of *Fusarium oxysporum* in liquid medium. *Toxicol Int.* 2011;18(2):111.
442. Matsakas L, Giannakou M, Vörös D. Effect of synthetic and natural media on lipid production from *Fusarium oxysporum*. *Electron J Biotechnol.* 2017;30:95–102.
443. Feofilova EP, Sergeeva YE, Mysyakina IS, Bokareva DA. Lipid composition in cell walls and in mycelial and spore cells of mycelial fungi. *Microbiology.* 2015;84(2):170–6.
444. Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, et al. MODOMICS: a database of RNA modification pathways—

- 2013 update. *Nucleic Acids Res.* 2012;41(D1):D262–7.
445. Yarian CS, Basti MM, Cain RJ, Ansari G, Guenther RH, Sochacka E, et al. Structural and functional roles of the N1-and N3-protons of Ψ at tRNA's position 39. *Nucleic Acids Res.* 1999;27(17):3543–9.
446. Sharma S, Watzinger P, Kötter P, Entian K-D. Identification of a novel methyltransferase, Bmt2, responsible for the N-1-methyl-adenosine base modification of 25S rRNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2013;41(10):5428–43.
447. Chan CTY, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, Begley TJ. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet.* 2010;6(12):e1001247.
448. Michielse CB, Reijnen L, Olivain C, Alabouvette C, Rep M. Degradation of aromatic compounds through the β -keto adipate pathway is required for pathogenicity of the tomato wilt pathogen *Fusarium oxysporum* f. sp. *lycopersici*. *Mol Plant Pathol.* 2012;13(9):1089–100.
449. Willke T, Vorlop K-D. Biotechnological production of itaconic acid. *Appl Microbiol Biotechnol.* 2001;56(3–4):289–95.
450. Okabe M, Lies D, Kanamasa S, Park EY. Biotechnological production of itaconic acid and its biosynthesis in *Aspergillus terreus*. *Appl Microbiol Biotechnol.* 2009;84(4):597–606.
451. Bentley R, Thiessen CP. Biosynthesis of itaconic acid in *Aspergillus terreus* I. Tracer studies with C14-labeled substrates. *J Biol Chem.* 1957;226(2):673–87.

452. Bonnarme P, Gillet B, Sepulchre AM, Role C, Beloeil JC, Ducrocq C. Itaconate biosynthesis in *Aspergillus terreus*. *J Bacteriol.* 1995;177(12):3573–8.
453. Michelucci A, Cordes T, Ghelfi J, Pailot A, Reiling N, Goldmann O, et al. Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *Proc Natl Acad Sci.* 2013;201218599.
454. Kubicek CP. The role of the citric acid cycle in fungal organic acid fermentations. In: *Biochemical Society Symposium.* 1987. p. 113–26.
455. Jaklitsch WM, Kubicek CP, Scrutton MC. The subcellular organization of itaconate biosynthesis in *Aspergillus terreus*. *Microbiology.* 1991;137(3):533–9.
456. Hama T, Ferré-D'Amaré AR. Pseudouridine synthases. *Chem Biol.* 2006;13(11):1125–35.
457. King RD, Lu C. An investigation into eukaryotic pseudouridine synthases. *J Bioinform Comput Biol.* 2014;12(4):1450015.
458. Anderson MZ, Brewer J, Singh U, Boothroyd JC. A pseudouridine synthase homologue is critical to cellular differentiation in *Toxoplasma gondii*. *Eukaryot Cell.* 2009;8(3):398–409.
459. Patton JR, Bykhovskaya Y, Mengesha E, Bertolotto C, Fischel-Ghodsian N. Mitochondrial myopathy and sideroblastic anemia (mlasa) missense mutation in the pseudouridine synthase 1 (*pus1*) gene is associated with the loss of trna pseudouridylation. *J Biol Chem.* 2005;280(20):19823–8.
460. Mason PJ, Wilson DB, Bessler M. Dyskeratosis congenita—a disease of dysfunctional telomere maintenance. *Curr Mol Med.* 2005;5(2):159–70.

461. Hynes MJ. Amide utilization in *Aspergillus nidulans*: evidence for a third amidase enzyme. *Microbiology*. 1975;91(1):99–109.
462. Fraser JA, Davis MA, Hynes MJ. The formamidase gene of *Aspergillus nidulans*: regulation by nitrogen metabolite repression and transcriptional interference by an overlapping upstream gene. *Genetics*. 2001;157(1):119–31.
463. Marzluf GA. Genetic regulation of nitrogen metabolism in the fungi. *Microbiol Mol Biol Rev*. 1997;61(1):17–32.
464. Talbot NJ, Ebbole DJ, Hamer JE. Identification and characterization of MPG1, a gene involved in pathogenicity from the rice blast fungus *Magnaporthe grisea*. *Plant Cell*. 1993;5(11):1575–90.
465. Van den Ackerveken G, Dunn RM, Cozijnsen AJ, Vossen J, Van den Broek HWJ, De Wit P. Nitrogen limitation induces expression of the avirulence gene *avr9* in the tomato pathogen *Cladosporium fulvum*. *Mol Gen Genet MGG*. 1994;243(3):277–85.
466. Coleman M, Henricot B, Arnau J, Oliver RP. Starvation-induced genes of the tomato pathogen *Cladosporium fulvum* are also induced during growth in planta. *Mol plant-microbe Interact*. 1997;10(9):1106–9.
467. Stephenson S-A, Green JR, Manners JM, Maclean DJ. Cloning and characterisation of glutamine synthetase from *Colletotrichum gloeosporioides* and demonstration of elevated expression during pathogenesis on *Stylosanthes guianensis*. *Curr Genet*. 1997;31(5):447–54.
468. Snoeijsers SS, Pérez-García A, Joosten MHAJ, De Wit PJGM. The effect of

nitrogen on disease development and gene expression in bacterial and fungal plant pathogens. *Eur J Plant Pathol.* 2000;106(6):493–506.

9. ANEXO 1

Artigo “A Computational Methodology to Overcome the Challenges Associated with the Search for Specific Enzyme Targets to Develop Drugs against *Leishmania major*.”

Reference: Catharina, L., Lima, C. R., Franca, A., Ramos Guimaraes, A. C., Alves-Ferreira, M., Tuffery, P., ... & Carels, N. (2017). A Computational Methodology to Overcome the Challenges associated with the Search for Specific Enzyme Targets to Develop Drugs against *Leishmania major*. *Bioinformatics and biology insights*, *11*, 1177932217712471.

10. ANEXO 2

Artigo “Specific enzyme functionalities of *Fusarium oxysporum* compared to host plants.”

Reference: Catharina, L., & Carels, N. (2018). Specific enzyme functionalities of *Fusarium oxysporum* compared to host plants. *Gene*, 676, 219-226.

11. ANEXO 3

Artigo “System Biology to Access Target Relevance in the Research and Development of Molecular Inhibitors.”

Reference: Catharina, L., de Menezes, M. A., & Carels, N. (2018). System Biology to Access Target Relevance in the Research and Development of Molecular Inhibitors. In *Theoretical and Applied Aspects of Systems Biology* (pp. 221-242). Springer, Cham.

A Computational Methodology to Overcome the Challenges Associated With the Search for Specific Enzyme Targets to Develop Drugs Against *Leishmania major*

Bioinformatics and Biology Insights
Volume 11: 1–21
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1177932217712471



Larissa Catharina¹, Carlyle Ribeiro Lima^{2,3}, Alexander Franca⁴, Ana Carolina Ramos Guimarães⁴, Marcelo Alves-Ferreira¹, Pierre Tuffery³, Philippe Derreumaux² and Nicolas Carels¹

¹Laboratório de Modelagem de Sistemas Biológicos, Instituto Nacional de Ciência e Tecnologia de Inovação em Doenças de Populações Negligenciadas (INCT-IDPN), Centro de Desenvolvimento Tecnológico em Saúde (CDTS), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, Brazil. ²Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique (UPR 9080), Centre National de la Recherche Scientifique (CNRS), Université Paris 7, Paris, France. ³Molécules Thérapeutiques in silico (UMR-S 973), Institut National de la Santé et de la Recherche Médicale (INSERM), Université Paris Diderot, Sorbonne Paris Cité, Paris, France. ⁴Laboratório de Genômica Funcional e Bioinformática, Instituto Oswaldo Cruz (IOC), Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, Brazil.

ABSTRACT: We present an approach for detecting enzymes that are specific of *Leishmania major* compared with *Homo sapiens* and provide targets that may assist research in drug development. This approach is based on traditional techniques of sequence homology comparison by similarity search and Markov modeling; it integrates the characterization of enzymatic functionality, secondary and tertiary protein structures, protein domain architecture, and metabolic environment. From 67 enzymes represented by 42 enzymatic activities classified by AnEnPi (Analogous Enzymes Pipeline) as specific for *L. major* compared with *H. sapiens*, only 40 (23 Enzyme Commission [EC] numbers) could actually be considered as strictly specific of *L. major* and 27 enzymes (19 EC numbers) were disregarded for having ambiguous homologies or analogies with *H. sapiens*. Among the 40 strictly specific enzymes, we identified sterol 24-C-methyltransferase, pyruvate phosphate dikinase, trypanothione synthetase, and RNA-editing ligase as 4 essential enzymes for *L. major* that may serve as targets for drug development.

KEYWORDS: Leishmaniasis, specific enzymes, sequence homology, metabolism, genomics, AnEnPi

RECEIVED: December 8, 2016. **ACCEPTED:** May 3, 2017.

PEER REVIEW: Six peer reviewers contributed to the peer review report. Reviewers' reports totaled 2767 words, excluding any confidential comments to the academic editor.

TYPE: Methodology

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is supported by fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (<http://www.capes.gov.br/>) to L.C.C., fellowship #11917-13-0 from Ciência sem Fronteiras (<http://www.cienciasemfronteiras.gov.br/web/csf/>) to C.R.L., Grant Universal 14/2013 #480866/2013-9 from Ministério de Ciência e Tecnologia/Conselho Nacional de

Desenvolvimento Científico e Tecnológico to A.C.R.G., and funding to N.C. under Instituto Nacional de Ciência e Tecnologia de Inovação em Doenças de Populações Negligenciadas #573642/2008-7.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Nicolas Carels, Laboratório de Modelagem de Sistemas Biológicos, Instituto Nacional de Ciência e Tecnologia de Inovação em Doenças de Populações Negligenciadas, Centro de Desenvolvimento Tecnológico em Saúde (CDTS), Fundação Oswaldo Cruz (Fiocruz), Av. Brasil 4036, 21040361 - RJ, Rio de Janeiro, Brazil. Email: nicolas.carels@cdis.fiocruz.br

Introduction

Leishmaniasis is a disease caused by a protozoan parasite from more than 20 species in the *Leishmania* genus of the family Trypanosomatidae that may affect skin (cutaneous form) and the mucosal membranes (visceral form); the visceral form can be fatal. Trypanosomatidae is successively inserted in the order Kinetoplastida, the subclass Metakinoplastina, the class Kinetoplastea, the phylum Euglenozoa and the clade Excavates. The parasite is transmitted by blood-sucking insects known as sandflies, a colloquial name for any species or genus of flying, biting, blood-sucking dipteran encountered in sandy areas. In the New World, leishmaniasis is spread by phlebotomine (Diptera, Psychodidae, Phlebotominae) of the genus *Lutzomyia*,¹ whereas in the Old World, it is spread by sandflies of the genus *Phlebotomus*.² The World Health Organization (WHO) estimates that there are 0.7 to 1.3 million new cases worldwide^{3,4} with 20 000 to 30 000 deaths per year. Medications include the following: (1) liposomal amphotericin B, a combination of

pentavalent antimonials and paromomycin, and miltefosine, for the visceral form and (2) paromomycin, fluconazole, or pentamidine, for the cutaneous form.⁵

Briefly, (1) liposomal amphotericin B is thought to act by binding to ergosterol, the principal sterol in fungal and *Leishmania* cell membranes,⁶ which results in a change in membrane permeability promoting monovalent ion leaks, metabolic disturbance, and cell death. AmBisome, the commercial formulation, is expensive, which makes its large-scale use in developing countries problematic. (2) The effect of pentavalent antimonials is pleiotropic because it may act (a) as a prodrug that is converted to active and more toxic trivalent antimony, (b) directly on molecular targets in the thiol redox metabolism⁷ such as trypanothione and glutathione,⁸ and (c) through thiols and ribonucleosides.⁹ (3) Paromomycin is an antibiotic that was proposed to successively (a) bind to the paraflagellar rod proteins and prohibitin, (b) be



internalized by endocytosis, and (c) interact with a P-type H⁺ adenosine triphosphatase (ATPase).¹⁰ (4) Miltefosine is an alkyl phospholipid whose mechanism of action is to inhibit phospholipid metabolism by decreasing phosphatidylcholine and increasing phosphatidylethanolamine¹¹ with the consequence of apoptosis-like cell death.¹²

Amphotericin B, pentavalent antimonials, paromomycin, and miltefosine induce noxious side effects and have variable efficacy in leishmaniasis treatment depending on the geographical locality. A combination treatment has the potential advantages of shortening the duration of treatment, reducing the overall dose of medicines, and reducing the probability of selection of drug-resistant parasites. Several trials of combinations have been conducted, with favorable results.¹³

In the case of the cutaneous form fluconazole and pentamidine may also be part of a combination.¹³ The New World form of *Leishmania* tends to be more severe and lasts longer than that of the Old World form. No single treatment approach fits all possible clinical presentations. In fungi, fluconazole interacts with 14- α demethylase, a cytochrome P450 enzyme necessary to convert lanosterol to ergosterol. As ergosterol is an essential component of the fungal cell membrane, inhibition of its synthesis results in increased cellular permeability, causing leakage of the cellular contents¹⁴; in *Leishmania*, the precise mechanism of action has not been described yet.¹⁵ Interestingly, it is well tolerated at relatively high doses by humans.¹⁶ Pentamidine was found to be a competitive inhibitor of arginine transport¹⁷ and a noncompetitive inhibitor of putrescine and spermidine transport in *Leishmania infantum*,¹⁸ *Leishmania donovani*, and *Leishmania mexicana*.¹⁹

Infections caused by *Leishmania* are becoming major public health problems on a global scale. Many species of *Leishmania* around the world are obtaining up to 15-fold resistance levels, as estimated by the WHO. The arsenal of drugs available for treating *Leishmania* infections is limited and includes pentavalent antimonials, pentamidine, amphotericin B, miltefosine, fluconazole, and a few other drugs at various stages of their development process.²⁰ *Leishmania* that is showing resistance is relatively difficult to observe and maintain in laboratory settings.²¹

The pharmaceutical industry has experienced a dramatic decrease in productivity between the 1980s and 2010 that is principally due to the cost burden of investing in the research and development (R&D) of new drugs,²² which are estimated to cost \$1.7 billion each.²³ In response, the process of drug R&D is shifting from how these activities were addressed as health care priorities in the past to approaches that are dominated by their potential market value. Even if the situation reverted to the figure of the 1980s (as defined by the number of new chemicals licensed by the Food and Drug Administration), concerns exist regarding future decision making, which requires a new paradigm for the management of R&D activities to attend to global needs.²⁴

Of course, screening for active compounds must be continued to anticipate the development of resistant parasitic forms. However, due to its low market potential, R&D for drugs against parasitic diseases that are endemic to tropical regions in developing countries is generally under prioritized or neglected by private companies and its realization depends on other means.^{25,26} As alternative financial sources for research on neglected diseases are limited and highly divided among individual researchers, *open science* and data sharing have received a growing interest as means of leveraging and combining the available resources to accelerate drug discovery efforts.²⁷ This community-based concept for a new drug discovery model led to the *London Declaration on Neglected Tropical Diseases* in 2012 for the control, elimination, or eradication of neglected tropical diseases (<http://unitingtofightntds.org>).

Many proteins are potential targets for drug interventions that control human diseases. The most recent number of drug targets was estimated to be in the hundreds, based on an analysis made before 2007.²⁸ However, the number of druggable proteins is substantially greater according to the DrugBank database Web site (<http://www.drugbank.ca/>). The current version of this database (5.0) contains 8206 drug entries that are linked to 4333 nonredundant (*nr*) protein sequences (ie, drug target/enzyme/transporter/carrier).

The discovery of drugs based on in silico docking of the inhibitors in models of the 3-dimensional (3D) structures of the protein targets has proven to be of great value in the processes of rational drug design as well as drug screening and has effectively contributed to conserving resources in the area of drug discovery. The use of this information may allow a substantial savings in the cost and time involved in the process of drug release.²⁹

Unfortunately, all druggable protein targets are not necessarily suitable for a therapeutic design because of several constraints, including noxious side effects to patients. The specific inactivation of a pathogenic enzyme without affecting any human enzymes would provide a safe approach to control the pathogen. The long-lasting, diverging evolution over a billion years that separates the common origins of simple unicellular and complex multicellular eukaryotes (phytozoan and metazoan) could have given rise to the enzymatic functions that are specific to parasitic pathogens in comparison with their hosts.³⁰ Similar mechanisms could generate analogous enzymes in pathogens and hosts that can also provide potential molecular targets due to the variations in their enzymatic sites that can be distinguished for inactivation by a given inhibitor.²⁹ By definition, analogous enzymes result from the convergent evolution of independent proteins rather than originating from common ancestral proteins (homologs), which allows for the same function but have differences in their primary, secondary, and tertiary structures.³¹

A previous study established a methodology to identify, annotate, compare, and study analogous and homologous

enzymes. The results obtained with this method could be used to identify the enzymatic activities in essential metabolic pathways that can serve as new therapeutic targets for fighting the infectious and parasitic diseases caused by protozoa of the *Trypanosoma* and *Leishmania* genera.³² This methodology includes the computational search for genes that encode (1) enzymes that are specific to parasites and, thus, are not encoded in human DNA, (2) analogous enzymes, and (3) the in silico mapping of the biological pathways of analogous and homologous enzymes considering the entire metabolism of the pathogens under consideration. The resulting pipeline, called AnEnPi (Analogous Enzymes Pipeline, <http://anenpi.fiocruz.br>), uses the protein sequences stored in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/>). AnEnPi implements algorithms that can perform the following tasks: (1) the clustering of protein sequences by homology using Basic Local Alignment Search Tool (BLAST); (2) the classification of sequences in different homologous clusters as analogous enzymes when they have the same enzymatic function, ie, Enzyme Commission (EC) number³³; (3) the detection of specific enzymes³⁴; (4) the annotation of protein functions using BLAST or HMMER; and (5) the generation of metabolic maps using the tools provided by KEGG (http://www.genome.jp/kegg/tool/map_pathway1.html). The reconstruction of metabolic pathways in parasites based on the metabolic maps provided by KEGG aims to identify the enzymatic activities that are essential for a parasite and can be considered promising targets for drug development.^{35–37}

Any error associated with a classification process can be qualified as follows: (1) *false positives*, which occur when a *Leishmania major* enzyme is classified as specific or analogous while at least 1 gene for a homologous enzyme exists in the *Homo sapiens* genome; (2) *false negatives*, which occur when the classification of a homologous pair is erroneous, and it should be classified as a specific or an analogous enzyme because of an incorrect annotation; (3) *true positives*, which occur when a pair of specific or analogous genes is correctly classified; and (4) *true negatives*, which occur when a pair of homologous genes indeed does not refer to specific or analogous genes.

The occurrence of *false positives* based on the available sequence annotations can result in significant errors and has been a recurring problem in the previously described methodology. Often, incomplete annotations are responsible for generating the *false positives* of specific or analogous enzymes. Because annotation is a dynamic process that improves with time, at any given moment, a pair of enzyme-encoding genes can be incorrectly classified as specific simply because the homologous counterpart in 1 genome of the pair has not yet been annotated.

Here, this report focuses on an evaluation of the *false positives* produced by AnEnPi because of the following: (1) in its current iteration, AnEnPi does not have any process of

false positive assessment and (2) this error component is critical for decisions on the investment of limited resources on a putative specific or analogous enzyme for drug development. The existence of *false positives* during target identification could invalidate years of research into drug development. We only considered the errors resulting from incomplete annotations here, not those from mistakes in the genome assemblies or EC annotations, and we describe a method to track *false positives* in the AnEnPi output to improve the identification of enzymes that are specific to *L major* compared with *H sapiens*.

As a result of the application of the proposed deputation process to *L major*, we found 4 specific enzymes (sterol 24-C-methyltransferase, RNA-editing ligase, pyruvate phosphate dikinase, and trypanothione synthetase) that seem to be valuable targets for further drug development.

Materials and Methods

Background

AnEnPi has produced *false positive* diagnoses due to inaccurate and fragmented annotations. Here, we propose a process to debug these error sources. We searched for *false positives* in a list of enzyme sequences obtained from AnEnPi³² whose functions were annotated as specific for *L major* compared with *H sapiens*. As a pipeline, AnEnPi classifies enzymes as homologous, analogous, and specific enzymes by considering the similarities among sequences with ≥ 100 amino acids and the functional annotations of the enzymes in the 2 organisms being compared. Considering all the combination pairs of enzymes between 2 organisms, an enzyme was designated (1) *homologous* to another when the similarity score and the E-value of their alignment was, respectively, ≥ 120 and close to 10^{-4} or below (using BLASTp); (2) *analogous* to another when the similarity score of their alignment was < 120 , but the enzymes were associated with the same EC (catalyzing the same reaction); (3) *functionally specific* when it was homologous to another, as defined under “1” but associated with a different EC; and (4) *strictly specific* when it was not homologous and did not share its EC with any of the other enzymes. In other words, by classifying enzymes as strictly specific to *L major*, we meant that the reactions of these enzymes were found to be catalyzed in *L major* but not in *H sapiens*. Thus, an enzyme was considered strictly specific to *L major* when its EC was not found in *H sapiens*. The gene list that we recovered from AnEnPi (KEGG release 58.1, June 1, 2011) included 67 sequences from *L major*, with 42 of these sequences being associated with a respective EC that was classified as strictly specific to *L major* compared with *H sapiens* by AnEnPi.

Process of strictly specific enzyme deputation

We could distinguish several steps and components in the process of strictly specific enzyme deputation. First, it is necessary to diagnose whether or not the putative strictly specific enzymes

of *L major* have homology with human DNA and its encoded enzymes. Second, a human subject enzyme that was found to be a homolog might be associated with the same EC as the *L major* query and could be diagnosed as a *false positive* or it could be associated with a different EC number and be diagnosed as functionally specific. Third, in the case of an ambiguous situation, we also characterized the 2-dimensional (2D) structure of the homologous regions. Fourth, we further checked putative strictly and functionally specific enzymes with 3D-HMM (HHpred and HHsearch) to detect eventual remote homologies in reference to the human 3D enzyme structures. Fifth, the enriched set of strictly specific enzymes of *L major* was challenged for the potentiality of its components as suitable as lead targets in reference to their centrality in metabolic pathways. Sixth, specific enzymes were investigated in other human parasites to assess their relevance as targets for drug development.

The process that we followed to enrich our enzyme set in the strictly specific proteins of *L major* that are potentially suitable for drug development is summarized in Figure 1 and is as follows. (1) The comparison of *L major* DNA-encoding proteins with human DNA was straightforward and it gave the sequence coordinates where homologous sequences could be found in the human genome (Figure 1A). The consistency of the detected homologies could be analyzed by performing multiple comparisons to determine whether the corresponding matches were obtained from protein-to-protein and protein-to-DNA using parasite-to-human and human-to-parasite queries (Figure 1B). (2) The amino acid sequence comparison between *L major* enzymes and human proteins allowed us to quickly identify most of the human proteins that deserved attention as potential homologous enzymes (Figure 1C). (3) The comparison of the human proteins identified under “2” with human chromosomal DNA allowed the delineation of their gene structure (exons and introns). (4) The comparison of the *L major* proteins that had a homologous hit with the human genomic sequence without a corresponding human protein in the Ensembl list³⁸ indicated a potentially missing human gene annotation or a possible pseudogene (Figure 1D). (5) The subject enzyme of a homologous pair is associated with an inconsistent annotation (Figure 1E) or a different EC compared with the query enzyme (Figure 1F). (6) We compared the 2D profiles of the query and subject sequences in ambiguous homologous pairs to discriminate orthology and distant paralogy or analogy (Figure 1G). (7) Ambiguous annotations, as in “6,” also occurred when an enzyme that was classified as non-homologous to the human genome was found to share an EC with that of a homologous one in *L major*, which is a case of intragenomic analogy. In that case, the nonhomologous *L major* enzyme might be analogous to the human counterpart of the homologous one, as it may occur when the EC of the human counterpart was incomplete or absent (Figure 1H). (8) We challenged the putative strictly and functionally specific enzymes of *L major* for remote homologies with human

enzymes using HHpred and HHsearch (Figure 1I). (9) We challenged the putative strictly specific enzymes of *L major* for their essentiality in KEGG metabolic pathways (Figure 1J). (10) To assess the relevance of the proposed strictly specific enzymes of *L major* for drug development, we analyzed the distribution of these enzymes in human parasites (Figure 1K).

Mapping Ensembl and L major homologous proteins on the human genome

To determine whether the classification of specific enzymes might produce *false positives*, we first searched for homology between the DNA-encoded proteins of *L major* and the DNA sequence of the human genome (Figure 1A). To this end, we employed the sequences from the putative *L major*-specific proteins available from TriTrypDB (release 6.0—September 2013)³⁹ and compared them with the human genome sequence (Ensembl, release 74—November 2013) using tBLASTn. We considered a pair to be a consistent homologous hit when the E-value of a tBLASTn alignment was $\leq 10^{-4}$ and its score value was ≥ 120 . When consistent hits were obtained, the genomic coordinates from the tBLASTn output were recovered and compared with the genomic coordinates of the encoded protein sequences available from Ensembl (release 74—November 2013) to determine whether a protein annotation might exist for the genomic region corresponding to the tBLASTn hit (Figure 2).

Thinking of the automation of that process, we identified 11 possibilities (Figure 2A), of which 7 are considered *TRUE* and 4 are considered *FALSE* in the Boolean sense. The *FALSE* and *TRUE* options can be easily diagnosed according to the decision tree shown in Figure 2B.

Consistency in homologous pair of L major vs H sapiens using BLAST

The lack of introns in *L major* genes may complicate the direct comparison of their gene sequences with those of the human genome, as the homologous regions, if any, could be interrupted by a splicing site in the human sequence of the homologous gene. To clarify this type of ambiguity, we checked whether a genomic hit obtained by comparing the *L major* proteins with the DNA sequence of human chromosomes (tBLASTn), as described in the previous paragraph, also matched the homologous region detected in the BLASTp (PSI-BLAST [Position-Specific Iterative BLAST]) comparison of the same *L major* query with the human protein sequence (subject) corresponding to that genomic hit. Thus, when an Ensembl protein annotation for a homologous hit of a *L major* protein with the human genome did exist (Figure 1B, Figure 3A), it was checked for consistency with the gene model, as found by tBLASTn comparison (Figure 3B) and from National Center for Biotechnology Information *consensus coding sequence* (CCDS)^{40,41} (Figure 3C), and, if consistent, it (them) was

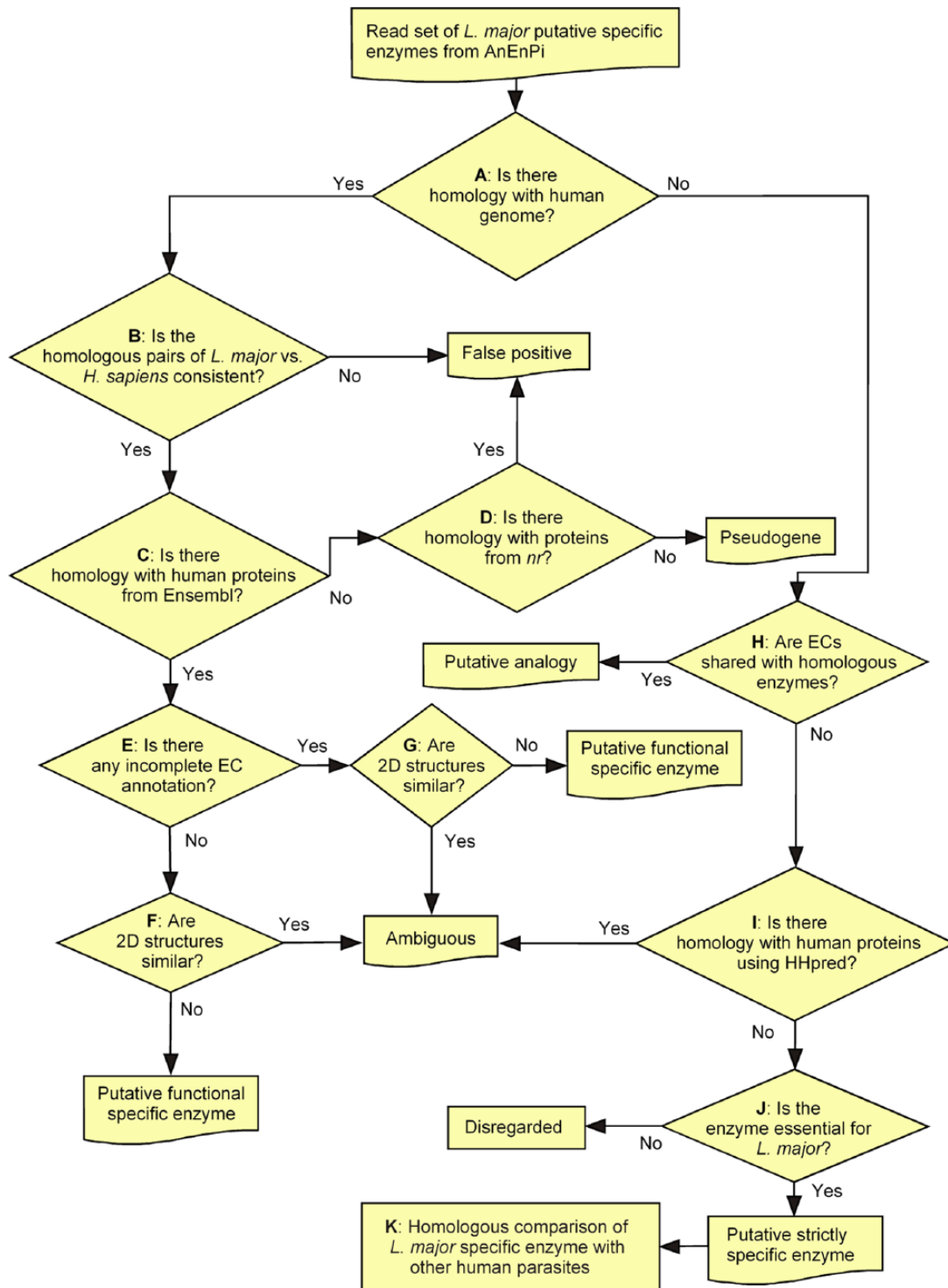


Figure 1. Flowchart of strictly specific enzyme depuration.

(were) assigned to the considered genomic region where its homology with the *L. major* enzyme was obtained.

To map a homologous region of a *L. major* protein on its tagged human gene (Figures 1A and 3), we must consider several lines of evidence. (1) Homology may be split into 2 homologous regions in the subject sequence if an intron falls in its genomic region corresponding to the protein domain of the query that is involved in the homology. In this case, the 2

subject homologies would be contiguous in the query regarding the final coordinate of the first region and the initial coordinate of the second region. (2) The genomic coordinate of the final position of a homologous region minus the genomic coordinate of its initial position plus 1 must be a multiple of 3 because of the codon structure of CCDSs. (3) The homologous region encodes a protein, and thus, the subject coordinates must match exons (Figures 1B and 3B).

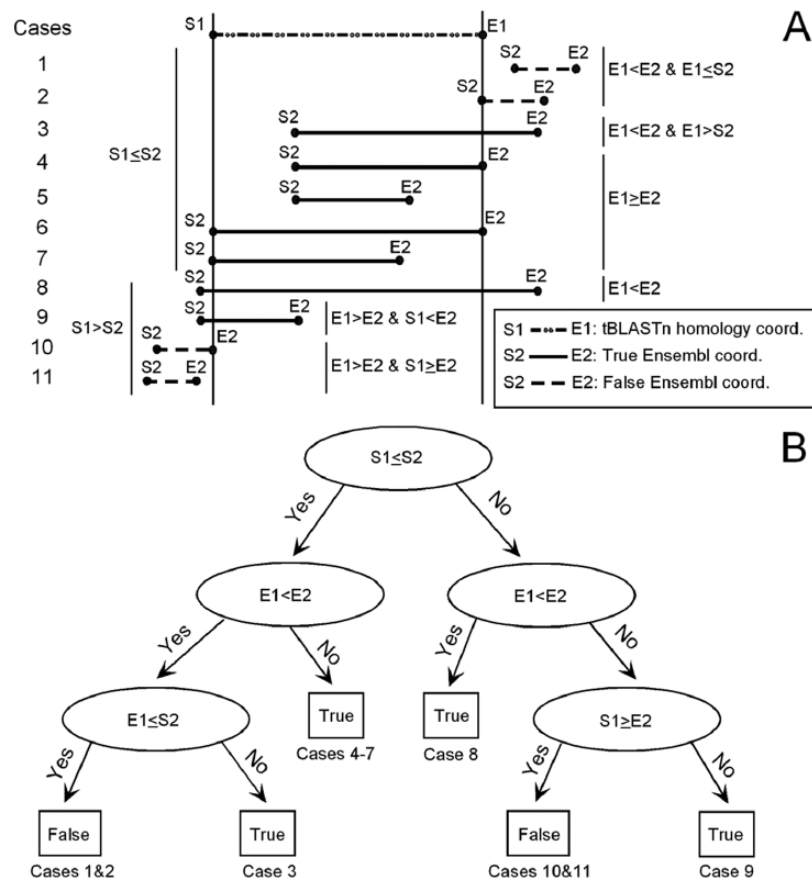


Figure 2. Homologous hit search algorithm for the classification of *false positives* of parasite-specific genes in humans. (A) The 11 possibilities of Ensembl protein associations (S2-E2) that one may obtain with a human genomic region that is homologous (tBLASTn) to a parasite protein query (S1-E1). S and E are for the beginning (start) and end of a tBLASTn homology or human gene coordinates. The Boolean description of each association between a tBLASTn hit in a human genome region and the Ensembl proteins annotated in that region is given on the left and right sides of panel A. “&” is used here in its Boolean sense, ie, a logical AND. Human genes for the Ensembl protein that are eventually compatible with a parasite’s homologous counterpart (tBLASTn) are modeled by thin lines. Human genes for Ensembl proteins in the same genomic region as a parasite’s homologous counterpart, but which are not compatible with it, are represented by dashed lines. (B): The decision tree for TRUE and FALSE associations of the human genes for the Ensembl proteins with a human genomic region that had a homology (tBLASTn) hit with a parasite protein.

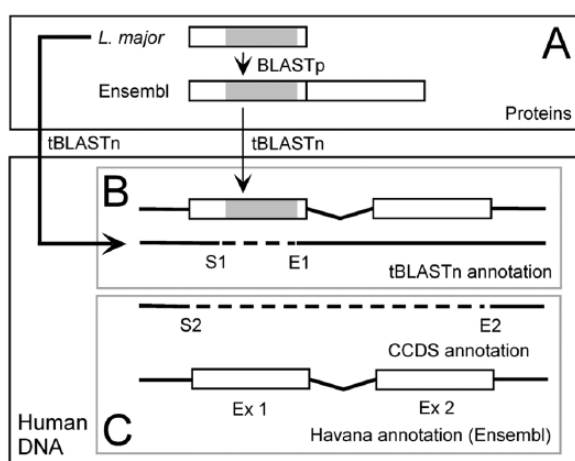


Figure 3. Search scheme for *false positives* of parasite-specific genes in human. The (A) homologous hit of *Leishmania major* and human proteins is compared with (B) the human gene structure obtained by tBLASTn search and with (C) the CCDS annotation from Havana.

After comparing the *L. major* proteins with human proteins (BLASTp) in Ensembl (Figures 1C and 3A), we may consider several steps to automate the process of homologous region mapping to human exons (Figure 3). (1) Considering the significant match of this comparison, we recovered the coordinates of the human gene according to the chromosome sequences (Figure 3B). (2) To determine the gene model corresponding to a human protein that has a significant hit with *L. major* proteins, we compared (tBLASTn) that human protein from Ensembl to its respective DNA sequence stretch by extracting it with the chromosomal coordinates of the corresponding initial and final gene position as informed by the Ensembl annotation (Figure 3C). (3) The homologies detected by the comparison of an Ensembl protein with its corresponding human DNA stretch, as described under “3,” gave us the coordinates for the exons (subject) relative to a gene model. (4) To translate the exon coordinates under “3” into coordinates relative to the chromosome context,

we added the exon coordinates to the initial position of the corresponding gene in chromosomal coordinates as provided by Ensembl. (5) To calculate the intron size, we subtracted the chromosomal coordinate (under “4”) of the beginning of an exon from the chromosomal coordinate of the end of the preceding exon for each exon of the corresponding gene. (6) To calculate the exon size, we subtracted the chromosomal coordinate (under “4”) for the end of an exon from the chromosomal coordinate for its beginning and added 1; we did this for all exons of each gene. Of course, the size of all the exons had to be a multiple of 3 (codons) to be considered relevant. (7) The sum of the exon size divided by 3 had to give us the protein size, which was inferred from the model obtained from the chromosomal sequence. The protein size could then be compared with the size of the query protein from Ensembl for size consistency. (8) Another consistency test was performed by determining whether the strands (“+” or “-”) from the exons, which were deduced from the subject homologies, were all “+” or all “-.” Exons in alternating “+” and “-” strands for the same gene with homology to the same query protein would not make sense, and this result allowed us to diagnose cases of inconsistency. (9) We also confirmed the gene models that we obtained from the tBLASTn search of the Ensembl proteins with human DNA by comparing the hit coordinates with the exon coordinates available in the CCDS database, as annotated by Ensembl. (10) To map the homologous region of a *L major* query with a human exon, we subtracted the initial and final chromosomal (human) coordinates (the subject) that corresponded to the hit with the initial and final coordinates of each human exon (obtained in “4”), respectively. The exon for which both subtractions gave values equal to 0 (or close to 0 providing that the small protruding end was a multiple of 3 bases) was considered the one that matched the homologous region in the *L major* query.

We considered the homologies between a *L major* protein and a human protein in the Ensembl list to be true positives when (1) the human protein was legitimate (all exons mapped on the same DNA strand) and (2) the homologous region between both proteins (from *H sapiens* and *L major*) matched the coding frame and coordinates of the genomic sequence that corresponded to the hit of the comparison (tBLASTn) of the *L major* proteins with the chromosomal DNA. In other words, to be considered significant, a homology between *L major* and a human protein had to be consistent with the gene model of that human protein, ie, the homologous regions between the proteins and DNA had to match the exons (Figures 1B and C and 3). All these operations were gathered together in a single Excel spreadsheet that we used as a dashboard (see Table S1).

Genomic hits without counterparts in Ensembl proteins

In cases where a genomic hit obtained by tBLASTn did not match the position of any human protein in the Ensembl list

(Figure 1D), it was considered as an indication that the human gene had not received a designation during the annotation process by Ensembl. The absence of an annotation was addressed by translating the homologous stretch of the human genome (subject) into amino acids and using the protein sequence for a homology search (BLASTp) with the *nr* section of GenBank (release 201.0—April 2014). When a human homologous protein was found in *nr*, its complete sequence was compared (tBLASTn) with a DNA stretch of 20 kb around the initial *L major* hit with the human genome to find the gene model (if any) associated with this protein.

The particular case of functional specificity

Considering the instances of enzymatic functional specificity between the 2 homologous proteins of *L major* and *H sapiens* in which the proteins were annotated as having differences in their enzymatic functionality (based on the assigned EC), we distinguished between the following categories (Figure 1E, F and H): (1) Uncertainty with the annotations occurred when at least one of the enzymatic functions in both proteins could not be confirmed (incomplete or unavailable EC). In these cases, we investigated the protein’s name and function in UniProtKB/SwissProt⁴² and attempted to verify whether the function described in this database matched that of TriTrypDB. If the function of the enzymes under comparison was synonymous in both organisms, the *L major* protein was diagnosed as a *false positive* for functional enzymatic specificity regarding its human counterpart; otherwise, functional specificity was assigned to the *L major* enzyme. This conservative position was taken because, if the members of a homologous pair had the same function, the potential of the negative collateral effects on the host of an inhibitor against the parasite’s enzyme would be greater if both the host and the parasite’s enzymes are orthologs. (2) Putative enzymatic functional specificity occurred in cases where the difference in enzymatic function of the homologous proteins could be confirmed. When both genes in the pair are paralogs rather than orthologs, they have an increased likelihood of having some sort of functional enzymatic specificity.¹¹ However, an inhibitor designed against the *L major* target may still affect the paralogous host counterpart at a lower rate.

Protein domain mapping and secondary structure

To better characterize whether ambiguous enzyme annotations (Figure 1G) were derived from distant paralogous or analogous enzymes, we searched their catalytic cores by mapping the domain composition for the conserved regions of the *H sapiens* and *L major* homologous pairs using PROSITE, Pfam,⁴³ and the Conserved Domain Database⁴⁴ for homology comparison with identity and score rates of ~40% and ≥ 120 , respectively, and predicted their secondary structure (2D) profiles using PRALINE.⁴⁵

Remote homology detection with hidden Markov models

In addition to BLAST, we used HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>)^{46,47} and HHsearch (<http://mobile.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py>),⁴⁸ which implement *hidden Markov models* (HMMs) for a pairwise comparison of the profiles from sequence alignments in databases (Figure 1I). HHpred detects the homologous sequences in humans with a higher sensitivity than BLAST. HHpred was performed with the default options, ie, local alignment and the scoring of secondary structure similarity, and HHsearch was executed with the Protein Data Bank (PDB; <http://www.rcsb.org/pdb/home/home.do>), HHblits (as the alignment generation method),⁴⁹ 85% (as sequence coverage), 10^{-6} (*P* value), and 10 (as maximal hit number). The sequences resulting from the first BLAST screening were further analyzed through HHpred and HHsearch to identify the possible remote sequence homologies in the human proteome that could result in *false positives* for the *L major*-specific enzymes. For a 3D model obtained by homology comparison to be consistent, a minimum identity level of 35% over 85% of the sequence template (subject) is required. Model consistency is necessary to evaluate the similarity between a query and its subject at the 3D level⁵⁰ to be able to diagnose whether both sequences are homologous or not. Alignments based on the primary sequences obtained from BLASTp may not have the necessary detection sensitivity and accuracy⁴⁹ to identify a homologous pair for the similarity levels at the border of the twilight zone.⁵¹ Thus, the comparatively higher sensitivity and accuracy of HHpred and HHsearch⁴⁶ is another filtering step that is necessary for detecting *false positives* in the putative strictly specific enzymes of *L major*.

Metabolic pathway significance

To clarify the relevance of the putative enzyme targets of *L major* in terms of the metabolic impairment that results from their inactivation, we searched their respective EC in the KEGG pathway database (<http://www.genome.jp/kegg/pathway.html>). Then, we searched their metabolic pathway insertion to diagnose whether their inactivation may potentially affect *L major* without having deleterious consequences on human metabolism and, consequently, whether these enzymes can be used as targets for drug development (Figure 1J).

Sequence comparison of strictly specific enzymes from *L major* with other parasites

Because an enzyme that was specific to *L major* compared with *H sapiens* could also have homologous pairs in other human parasites (Figure 1K) and eventually serve as a target for drug development in these organisms, we searched for sequence homologies between the strictly specific enzymes of *L major* and the proteins in the organisms listed in Table 1.

When an enzyme from these organisms was found to be present in KEGG (release 81.0—January 1, 2017) by comparing their EC annotation with that of *L major*, we acknowledged it. When it was not present in an annotation comparison, we proceeded with a sequence comparison using BLASTp. We acknowledged the homologies with the *L major* query as significant when the homology region was at least 75% of the query and the identity was $\geq 35\%$, had an E-value $\leq 10^{-4}$, and a score ≥ 120 .⁵⁰

Phylogenetic analysis

On one hand, the lethal effect of REL1 inhibition has been shown in *Trypanosoma brucei* by Panigrahi et al,⁵² who identified MP52 as the sequence associated with REL1.⁵³ The association of REL1 sequences between *T brucei*, *Trypanosoma cruzi*, and *L major* has been clearly established by alignment⁵⁴ and allowed us to unambiguously assign the REL1 function to LmjF.01.0590. On the other hand, complete genome sequences were produced for *Bodo saltans* and *Trypanoplasma borreli* that belong, respectively, to 2 different orders, ie, Eubodonida and Parabodonida, within Metakinetoplastina.⁵⁵ This relation allowed us to test the hypothesis of whether the REL sequences of EC 6.5.1.3 emerged only within Trypanosomatidae or at the higher taxonomical level of Metakinetoplastina.

From the complete genome sequences of *B saltans* (GenBank Assembly ID: GCA_001460835.1, release ID: 2707578) and *T borreli* (<http://www.sanger.ac.uk/resources/downloads/protozoa/trypanoplasma-borreli.html>), we found the sequences encoding REL1 within the sequence NODE_50574 through a tBLASTn search of LmjF.01.0590 in the file *Trypanoplasma borreli*_contigs_20120411.fa.gz.

We aligned the REL1 sequences of *T brucei*, *T cruzi*, *L major*, *B saltans*, and *T borreli* with ClustalW⁵⁶ and eliminated the column corresponding to gaps, and we searched the best phylogenetic relationship using maximum likelihood with MEGA7.⁵⁷ The substitution model was *Jones–Taylor–Thornton*, the rate among sites was considered uniform, and the tree inference was obtained by *nearest neighbor interchange*.

Results

Filtering pipeline of strictly specific enzymes of *L major* compared with *H sapiens*

Comparing the complete sets of (1) 8265 sequences of *L major* with 1547 that are annotated with ECs (341 *nr* ECs) and (2) 20834 sequences of *H sapiens* with 6599 that are annotated with ECs (384 *nr* ECs) from KEGG, AnEnPi released a list of 67 protein sequences corresponding to 42 enzymatic activities that were putatively specific to *L major* compared with humans.

From the 67 enzymes (42 ECs) released by AnEnPi, 15 sequences (13 ECs) from *L major* produced 79 homologous hits with the human genome in CCDSs, as annotated by Ensembl, leaving 52 *L major* enzymes (29 ECs) without any

Table 1. Sequence materials of some important human parasites for comparison with the strictly specific enzymes of *Leishmania major*.

ORGANISM	GENBANK ASSEMBLY ID ^a	RELEASE ID	REFSEQ ASSEMBLY ID ^a	RELEASE ID
Fungi				
<i>Cryptococcus neoformans</i>				
Serotype JEC 21	091045.1	111678	091045.1	256138
Serotype B-3501 A	149385.1	172358	149385.1	172378
<i>Histoplasma capsulatum</i>				
Serotype Nam1	149585.1	172628	149585.1	293588
<i>Coccidioides immitis</i>	149335.2	292488	149335.2	1741148
Helminths				
<i>Schistosoma mansoni</i>				
	237925.2	557858	237925.1	3549348
<i>Brugia malayi</i>				
	002995.2	449758	002995.3	451858
<i>Loa loa</i>				
	183805.3	1163928	183805.2	4196238
Protozoa				
<i>Toxoplasma gondii</i>				
	006565.2	751938	006565.2	3557728
<i>Trypanosoma cruzi</i>				
	209065.1	271488	209065.1	271508
<i>Trypanosoma brucei</i>				
	002445.1	522118	002445.1	2888
<i>L. major</i>				
	002725.2	283598	002725.2	402998
<i>Leishmania donovani</i>				
Strain BPK282A1	227135.2	405018	227135.1	405068
<i>Leishmania braziliensis</i>				
Strain M2904	002845.2	515698	002845.2	515718
<i>Plasmodium falciparum</i>				
Serotype 3D7	002765.1	256198	002765.3	360518
Serotype Dd2	149795.1	172858	NA	NA
Serotype HB3	149665.2	2204658	NA	NA
<i>Giardia lamblia</i>				
	002435.1	278778	002435.1	284418
Bacteria				
<i>Staphylococcus aureus</i> JH9 (MRSA/VRSA)				
	016805.1	17608	016805.1	37908
<i>Streptococcus pyogenes</i> M1 GAS				
	006785.2	1009368	006785.2	1851728
<i>Mycobacterium tuberculosis</i>				
Strain H37Rv	195955.2	538028	195955.2	538048

Abbreviations: GAS, group A streptococci; MRSA, methicillin-resistant *S. aureus*; NA, not assigned; VRSA, vancomycin-resistant *S. aureus*.

^aAssembly ID in the format XXX_000YYYYYY.Z where X is the Bank ID number such that GCA and GCF indicate a sequence from GenBank and RefSeq, respectively, Y indicates the accession number, and Z is the version number.

homologies in the human genome (Figure 4A). By comparing the coordinates of the gene structures (exons + introns) reported by Ensembl for these CCDSs with the coordinates of the homologies given by the tBLASTn search of their protein

sequence with the human genome (putative exons), we identified many inconsistencies in the Ensembl CCDS annotations (Figure 4B). By analyzing the 79 hits, we identified 64 nonconsistent CCDSs where the putative exons were (1) associated

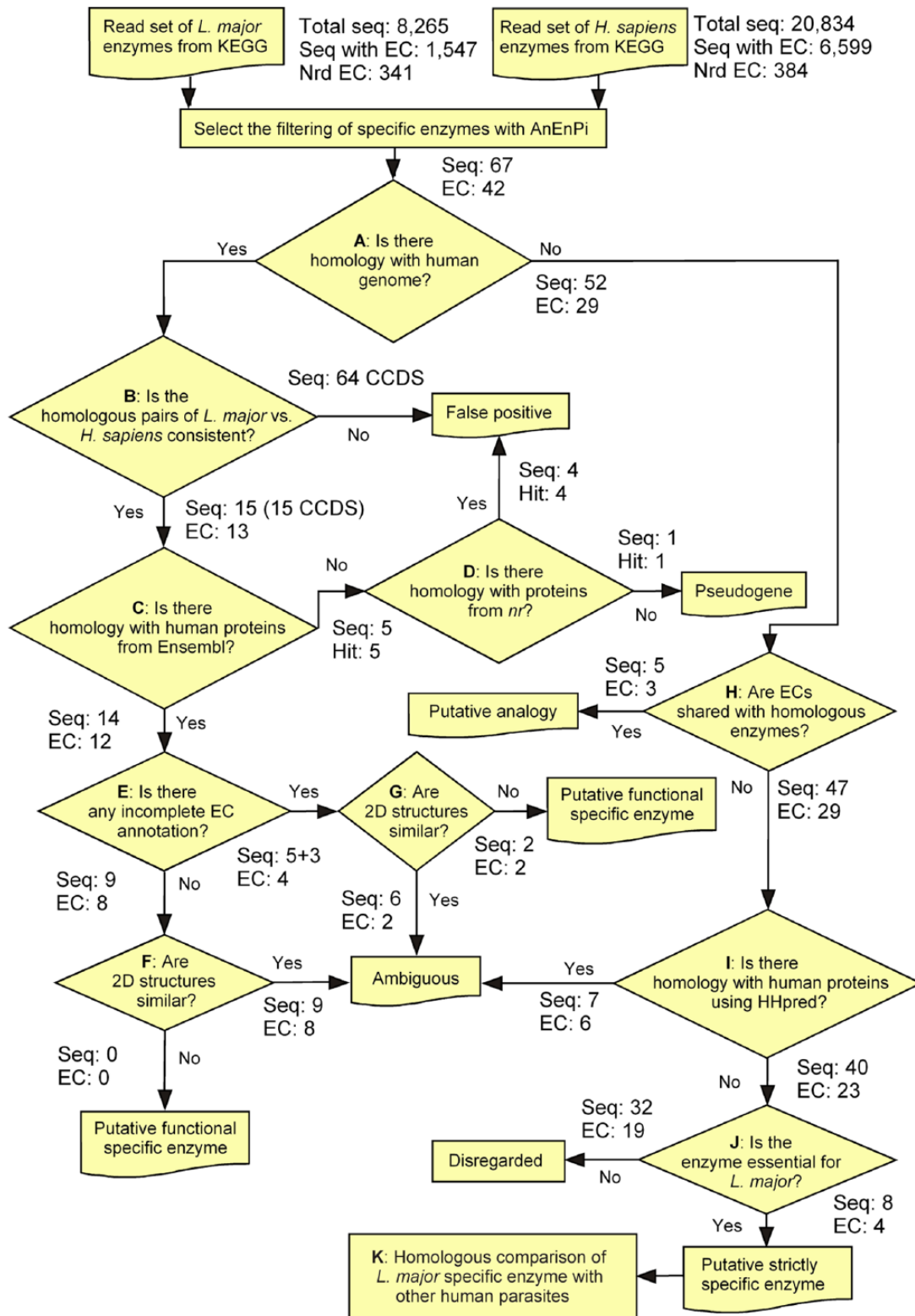


Figure 4. Flowchart of the sequence depuration process in *Leishmania major* compared with *Homo sapiens*. Nrd stands for nonredundant.

with the same protein, but they appeared on both gene strands, which is impossible given the translation process or (2) simply deprived of a protein association, making any further inference impossible (see Table S1).

Among the 15 *L. major* enzymes, 14 demonstrated homology to the *H. sapiens* proteins from Ensembl and 1

(LmjF.23.0270, pteridine reductase 1, EC 1.5.1.33) had a hit with the human genome sequence itself, but not with the CCDSs annotated by Ensembl (Table 2). Table 2 shows that in addition to LmjF.23.0270, 4 other *L. major* enzymes also had hits with the human genome, but that they were not annotated as CCDS by Ensembl. However, in these 4 cases (LmjF.18.1510,

Table 2. *Homo sapiens* proteins retrieved from *nr* using the DNA stretch of the subject, which corresponded to the homologous region between the *Leishmania major* proteins and the human genome, as a query.

<i>L MAJOR</i> ID	CHR. NO. ^a	HIT, GENOMIC COORDINATES	HUMAN HOMOLOGOUS PROTEINS IN NR	
			FUNCTION	ACC. NO. ^b
LmjF.18.1510, LmjF.18.1520	8	19098173-19098421	PMCA2a, partial	NP_064524 ^c
LmjF.18.1510, LmjF.18.1520	18	19998852-19999097	hCG2039601	EAX0113 ^d
LmjF.23.0270	6	99622932-99623354	3-hydroxybutyrate dehydrogenase, type 2	AAH01953 ^c
LmjF.27.2440	6	99622770-99623351	3-hydroxybutyrate dehydrogenase, type 2	ABG2424 ^c
LmjF.26.2280	1	78509203-78509814	NA	NA

Abbreviation: NA, not assigned.

^aChr. no. is for chromosome number.

^bAcc. no. is for accession number.

^cPseudogenes.

^dMissing annotation.

LmjF.18.1520, LmjF.26.2280, and LmjF.27.2440), homologues were also found with CCDSs at other genomic coordinates. In *nr*, we found homologous protein sequences for the human genomic hits corresponding to LmjF.23.0270, LmjF.18.1510, LmjF.18.1520, and LmjF.27.2440 that were not annotated as CCDS by Ensembl. We could retrieve a complete gene model only for EAX0113, which meant that the protein hCG2039601 between 19998852 and 19999097bp on chromosome 18 escaped an Ensembl annotation. The gene models for the 3 homologues at human chromosomes 8 and 6 were partial, which indicates that they should likely be considered pseudogenes. Similarly, LmjF.26.2280 (EC 3.5.5.1—nitrilase) matched a pseudogene because an inframe stop codon (“*”) could be found in the sequence of the human homologous region (subject) in the tBLASTn alignment (Figure 4D).

The equivalent KEGG accession and ECs of the 17 human proteins from Ensembl that were homologous to the 14 *L major* enzymes (14 ECs and 3 repeats), given in Table 3, could be classified into 3 categories (Figure 4C). (1) Nine enzymes were annotated with 8 different ECs (1 repeat) in *L major*, which means that they were associated with different enzymatic reactions in *L major* and humans, ie, they could be considered specific from a functional standpoint. (2) Three *L major* enzymes were associated with 2 incomplete ECs in humans (1 repeat). (3) Five (1 repeat with the category “different EC numbers”) were annotated as enzymes in *L major* (4 ECs), but their human homologs were not annotated as enzymes (undef.) (see Table 3). It is worth noting here that in all the 17 cases, the homologous region between the *L major* protein and the human genome sequences precisely corresponded to a unique exon. These conserved regions, corresponding to the protein domains, were never interrupted by an intron in the human genome.

By comparing the α -helix and β -sheet distributions in the homologous pairs under “different EC number” and “incomplete

EC number” (Table 3), we did not find any significant difference in the 2D profiles, which suggested that the classification of these sequences as functionally specific was ambiguous. It is only in the 5 sequences (4 ECs) under “EC number not assigned” (Table 3) that we found significant differences in the 2D profiles between both regions of the homologous pairs (Table 3, Figure 4F). However, among these 5 enzymes, only 2 (LmjF.14.0350 and LmjF.30.0180) were not involved in homologies with those of “different EC number.” These 2 enzymes in *L major* were composed of more β -sheets than α -helices, which did not occur in their human counterparts where the ratio between β -sheets and α -helices was approximately the same.

The relation of 2D distribution analysis between the *L major* and human conserved domains led us to finally consider 2 sequences from the 15 sequences (13 ECs) that were homologous to humans as being functionally specific enzymes in *L major* compared with *H sapiens*.

Considering the putative strictly specific enzymes of *L major* that shared ECs with the sequences classified as putative functionally specific enzymes, we found that 5 sequences (LmjF.18.0560, LmjF.19.1020, LmjF.26.0420, LmjF.36.1660, and LmjF.28.2100) shared 3 ECs (Table 4). Because these sequences are strictly specific, they are not expected to have human counterparts; however, the fact that they share ECs with the group of enzymes that are homologous to the human genome implies that they could be analogous to them. We also disregarded these 5 sequences, as we were only interested in filtering strictly specific enzymes in this report.

Strictly specific enzymes of L major

By further checking for remote homology with HHpred and HHsearch (Figure 4I), 7 sequences (LmjF.07.0270, LmjF.14.0180, LmjF.16.0530, LmjF.17.0140, LmjF.18.0200, LmjF.33.2540, and LmjF.36.3590) (6 ECs) were considered distant homologs

Table 3. EC number annotations of the human proteins that are homologous to the *Leishmania major* enzymes classified as functionally specific.

ACCESSIONS	EC NO.	ENZYMATIC FUNCTION
Different EC numbers		
LmjF.36.5960 ^a	3.1.4.46	Glycerophosphodiester phosphodiesterase
Hsa:51573 ^b /006007 ^c	3.1.4.44	Glycerophosphoinositol glycerophosphodiesterase
LmjF.35.4250	2.1.1.64	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase
Hsa:51805/132423	2.1.1.114	Polyprenyldihydroxybenzoate methyltransferase
LmjF.28.2910	1.4.1.4	Glutamate dehydrogenase (NADP+)
Hsa:2747/182890	1.4.1.3	Glutamate dehydrogenase [NAD(P)+]
LmjF.18.1510	3.6.3.6	H ⁺ -exporting ATPase
Hsa:478/105409	3.6.3.9	Na ⁺ /K ⁺ -exchanging ATPase
LmjF.18.1520	3.6.3.6	H ⁺ -exporting ATPase
Hsa:478/105409	3.6.3.9	Na ⁺ /K ⁺ -exchanging ATPase
LmjF.33.0680	1.3.1.71	Delta24(241)-sterol reductase
Hsa:7108/149809	1.3.1.70	Delta14-sterol reductase
LmjF.05.0350	1.8.1.12	Trypanothione reductase
Hsa:7296/198431	1.8.1.9	Thioredoxin reductase (NADPH)
LmjF.09.0770	3.4.21.83	Oligopeptidase B
Hsa:5550/085377	3.4.21.26	Prolyl oligopeptidase
LmjF.30.1550	4.2.1.70	Pseudouridylate synthase
Hsa:83480/110060	5.4.99.45	tRNA pseudouridine (38/39) synthase
Incomplete EC number		
LmjF.26.2280	3.5.5.1	Nitrilase
Hsa:4817/158793	3.5.-.-	Nitrilase 1
LmjF.24.2030	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
Hsa:84869/145439	1.1.1.-	Oxidoreductase
LmjF.27.2440	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
Hsa:84869/145439	1.1.1.-	Oxidoreductase
EC number not assigned		
LmjF.14.0350	4.2.3.1	Threonine synthase
Hsa: NA/185875	-.-.-	
LmjF.30.1550	4.2.1.70	Pseudouridylate synthase
Hsa: NA/198331	-.-.-	
LmjF.30.0180	1.1.1.60	2-hydroxy-3-oxopropionate reductase
Hsa: NA/140632	-.-.-	
LmjF.18.1510, LmjF.18.1520	3.6.3.6	H ⁺ -exporting ATPase
Hsa: NA/123191	-.-.-	

Abbreviations: ATPase, adenosine triphosphatase; EC, Enzyme Commission; NA, not assigned; tRNA, transfer RNA.

^aKEGG accession numbers for *L. major*.

^bKEGG accession numbers for *Homo sapiens*.

^cEnsembl accession numbers for *H. sapiens* according to the format ENSG00000xxxxxx.

Table 4. Putative functional and strictly specific enzymes of *Leishmania major* that share ECs.

<i>L MAJOR</i>			<i>HOMO SAPIENS</i> ^a		PAIRING ^d
EC	SPECIFICITY		EC ^b	ACC. ^c	
	FUNCTIONAL	STRICT			
3.6.3.6	LmjF.18.1510	—	3.6.3.9/NA	105409/123191	Homologous
3.6.3.6	LmjF.18.1520	—	3.6.3.9/NA	105409/123191	Homologous
3.6.3.6	—	LmjF.18.0560	—	—	Analogous
4.2.1.70	LmjF.30.1550	—	5.4.99.45	110060	Homologous
4.2.1.70	—	LmjF.19.1020	—	—	Analogous
4.2.1.70	—	LmjF.26.0420	—	—	Analogous
4.2.1.70	—	LmjF.36.1660	—	—	Analogous
3.1.4.46	LmjF.36.5960	—	3.1.4.46	006007	Homologous
3.1.4.46	—	LmjF.28.2100	—	—	Analogous

Abbreviation: EC, Enzyme Commission; NA, not assigned.

^aECs and Ensembl accession number for human sequences homologous with the putative functional specific enzymes of *L major*.

^bWhen a human EC was not available under “EC number not assigned,” we gave the one available under “different EC numbers” (Table 3) following the pattern “different EC numbers”/“EC number not assigned.”

^cWe applied the same pattern as in “b” for Ensembl accession numbers.

^dWhen an Ensembl protein is homologous to a putative functional specific enzyme of *L major* is available, their pairing relationship must be about homology; when such an Ensembl protein is not available, there is no homology and the pairing is only about enzymatic function (EC), ie, analogy.

and were filtered out from the putative strictly specific enzymes approved under Figure 4H (47 sequences, 29 ECs).

A homology analysis of the remaining 40 sequences (23 ECs) (Table 5) among Trypanosoma and *Plasmodium falciparum* shows that *L major* (query) shares the following: (1) 9 sequences with 24 sequences from *T cruzi*, (2) 7 sequences with 18 sequences from *T brucei*, and (3) 2 sequences with 1 sequence from *P falciparum* (Table S2).

From the 40 strictly specific enzymes (23 ECs) (Table 5), 32 (19 ECs) were further disregarded because they did not seem to match any essential enzymes with a central position in the metabolic maps, as documented by KEGG (Figure 4J).

Only 8 specific enzymes, corresponding to 4 ECs (EC 2.1.1.41 with 2 paralogous sequences: LmjF.36.2380 and LmjF.36.2390; EC 2.7.9.1 with 1 sequence: LmjF.11.1000; EC 6.3.1.9 with 3 paralogous sequences: LmjF.23.0460, LmjF.27.1870, and LmjF.36.4300; and EC 6.5.1.3 with 2 paralogous sequences: LmjF.01.0590 and LmjF.20.1730), could be considered as potential target candidates for drug development. The complete relation of gene paralogy, analogy, and uniqueness of the 67 sequences released by AnEnPi as putative specific enzymes of *L major* compared with human is given in Table S3.

Evolutionary relationship of key strictly specific enzymes in *L major* with other human parasites

To better understand the consequences of the evolutionary relationship of key strictly specific enzymes for drug development (Figure 4K), we searched for homologous proteins in other

human parasites (Table 6). It appeared that EC 2.1.1.41 (sterol 24-C-methyltransferase) is conserved among fungi and *L major*, which is expected from their common use of the ergosterol pathway. Interestingly, the homologies for EC 2.7.9.1 (pyruvate phosphate dikinase) are only shared among protozoa; the enzyme from *Mycobacterium tuberculosis* must be considered analogous because it does not share homologies with the *L major* sequences. As can be deduced from Table 6, EC 6.3.1.9 (trypanothione synthetase) is strictly specific to Trypanosoma, which confirms previous investigations. At first glance, EC 6.5.1.3 (mitochondrial RNA-editing ligase) is a function that appears relatively well conserved among human parasites. However, REL1 (LmjF.01.0590) and REL2 (LmjF.20.1730) are not homologous to the transfer RNA (tRNA)-splicing ligase (RtcB) of *P falciparum* and *Toxoplasma gondii* (Apicomplexa), even if they were annotated with the same enzymatic function (EC 6.5.1.3). Thus, the RELs from Trypanosomatidae and RtcBs from Apicomplexa should be considered analogous. The phylogenetic tree in Figure 5 shows that REL1 is conserved into Metakinetoplastina as far as *B saltans* and *T borreli* are representatives of this subclass, together with Trypanosomatidae. REL2 has not been taken into consideration here because it was proven to be nonessential in *T brucei*.⁵²

Discussion

Classification challenges and annotation ambiguity

In this study, we described a process for filtering out the false-positive annotations of enzymatic specificity in the host-parasite

Table 5. Strictly specific enzymes of *Leishmania major* after hidden Markov model filtering.

ACC.	ENZYMATIC FUNCTION	EC	METABOLISM
LmjF.07.0260	Homoserine dehydrogenase-like protein	1.1.1.3 ^a	Amino acids
LmjF.17.1360	L-galactonolactone oxidase	1.1.3.8 ^a	Ascorbate
LmjF.34.0070	Putative ascorbate peroxidase	1.11.1.11	Ascorbate
LmjF.26.2610	FAD-binding domain containing protein, putative	1.14.13.1 ^a	Polycyclic degradation
LmjF.30.0610	Nitrate reductase, putative	1.7.1.1 ^a	Nitrogen
LmjF.31.0010	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase	2.1.1.14 ^a	Amino acids
LmjF.36.2380, LmjF.36.2390	Sterol 24-C-methyltransferase, putative	2.1.1.41	Steroids
LmjF.09.1040	Phospholipid:diacylglycerol acyltransferase, putative	2.3.1.158 ^a	Glycolipids
LmjF.30.3080	Homoserine kinase	2.7.1.39	Energy, amino acids
LmjF.17.1160	UDP-sugar pyrophosphorylase	2.7.7.64	Carbo. Ascorb.
LmjF.11.1000	Pyruvate phosphate dikinase, putative	2.7.9.1	
LmjF.36.4640	Esterase PHB depolymerase, putative	3.1.1.73 ^a	Carbohydrate
LmjF.31.2300, LmjF.31.2310	3'-nucleotidase/nuclease, putative	3.1.30.1	Nucleic acids
LmjF.04.0310, LmjF.04.0320, LmjF.23.0870, LmjF.23.0880, LmjF.35.0640	β -Fructofuranosidase, putative	3.2.1.26	Carbohydrate
LmjF.29.2800	Inosine-adenosine-guanosine-nucleosidehydro-lase, putative	3.2.2.1	Nucleotide
LmjF.10.0460, LmjF.10.0465, LmjF.10.0470, LmjF.10.0480, LmjF.31.2000, LmjF.28.0570	GP63, leishmanolysin (protease) ³³	3.4.24.36	
LmjF.34.3250	α -Keto-acid decarboxylase, putative	4.1.1.74	Amino acids
LmjF.09.0360, LmjF.33.0470	DNA photolyase, putative	4.1.99.3	Nucleotide
LmjF.14.0460	Cystathionine β -lyase-like protein	4.4.1.8	Amino acids
LmjF.19.0985, LmjF.19.0995, LmjF.19.1005	4-coumarate:coA ligase-like protein	6.2.1.12 ^a	Amino acids
LmjF.26.0830	Aspartate-ammonia ligase, putative	6.3.1.1	Amino acids
LmjF.27.1870, LmjF.36.4300, LmjF.23.0460	Trypanothione synthetase, putative	6.3.1.9	Glutathione
LmjF.01.0590, LmjF.20.1730	Mitochondrial RNA-editing ligase	6.5.1.3	Nucleic acids

Abbreviations: EC, Enzyme Commission; FAD, flavin adenine dinucleotide; UDP, uridine diphosphate.

^aEC numbers without any previous report in the Tritryps literature for drug development.

interactions released by AnEnPi.³² The identification of specific protein targets in parasites (here, *L major*) compared with their host (here, *H sapiens*) is important for the development of drugs with the least amount of negative collateral effects for the host as possible. The procedure presented here is systematic, compatible with automation and suitable for host-parasite interactions that involve a lower eukaryote as the parasite and a higher eukaryote as the host. The fact that the targets identified here have been previously identified suggests that the results are valid and increases our confidence that the proposed methodology can be applied to other host-parasite systems with a lower level of associated knowledge.

We analyzed 67 protein sequences that correspond to 42 enzymatic activities and were previously classified as specific to *L major* compared with *H sapiens* by AnEnPi. A careful homology comparison of these sequences with those of Ensembl using BLAST allowed us to identify 15 significant homologous pairs between *L major* and *H sapiens*. These homologous pairs were associated with different ECs in *L major* and *H sapiens*, suggesting that they should be classified as functionally specific in *L major* compared with *H sapiens*. Of course, the classification of enzymes with functional specificity is completely dependent on the EC annotations, and it may result in trivial errors, as is the case for the ATPases that

Table 6. Distribution of strictly specific enzymes in human parasites.

ORGANISMS	EC NUMBERS			
	2.1.1.41	2.7.9.1	6.3.1.9	6.5.1.3
Fungi				
<i>Cryptococcus neoformans</i>	XP_777275.1 ^a , XP_568887.1	—	—	A ^b
<i>Histoplasma capsulatum</i>	XP_001537255.1	—	—	A
<i>Coccidioides immitis</i>	XP_001242243.2, XP_001241281.1	—	—	A
Helminths				
<i>Schistosoma mansoni</i>	—	—	—	A
<i>Brugia malayi</i>	—	—	—	A
<i>Loa loa</i>	—	—	—	A
Protozoan				
<i>Toxoplasma gondii</i>	—	—	—	A
<i>Plasmodium falciparum</i>	—	—	—	A
<i>Trypanosoma cruzi</i>	X ^c	X	X	X
<i>Trypanosoma brucei</i>	X	X	X	X
<i>Leishmania donovani</i>	X	X	X	X
<i>Leishmania braziliensis</i>	X	X	X	X
<i>Giardia lamblia</i>	—	XP_001705572.1	—	—
Bacteria				
<i>Staphylococcus aureus</i>	—	—	—	—
<i>Streptococcus pyogenes</i>	—	—	—	—
<i>Mycobacterium tuberculosis</i>	—	A	—	A

Abbreviation: EC, Enzyme Commission.

^aAccession number for homologous sequences.

^bA is for analogous sequences, when the homology could not be confirmed but the EC was annotated in KEGG.

^cX is for homologies among Tritryps; the details of these homologies if given in Table S3.1.

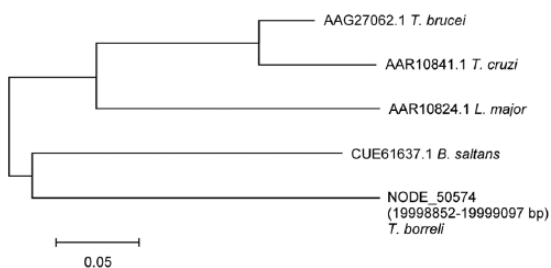


Figure 5. Phylogenetic tree of REL1 in Metakinetoplastina as obtained by maximum likelihood. Scale bar: 0.05 substitutions per site.

may carry H⁺ or Na⁺ cations through cell membranes. The 2 ATPases carrying H⁺ or Na⁺ are homologous, and the difference in their EC numbers is associated with the transported cations (the reaction and not the enzymatic activity), which are not associated with the ATPase activity itself when the substrate is adenosine triphosphate (ATP). Thus, care must

be taken when interpreting ECs. In any event, we only found minute differences when considering the profiles of 2D structures between the sequence members of the BLAST homologous pairs.

Among the 15 putative functionally specific enzymes, the ECs of 9 *L. major* proteins only differed from those of *H. sapiens* in the fourth digit. When comparing the 2D alignments of these 9 homologous pairs, we only found slight differences, and it is difficult to claim that a drug for these *L. major* targets would not have some kind of unwanted interaction with the human form that could result in negative side effects for the patients.

The Ensembl database involves processes of automatic gene annotation that are eventually manually curated; this is also the case for Havana.^{38,58} The AnEnPi pipeline is completely automatic, and it classifies enzymes as specific, homolog, or analog according to their homology and

enzymatic activities. Obviously, incomplete enzymatic annotations lead to nonoptimized classifications by AnEnPi. The purpose of this study was to detect these cases and to provide more knowledge for improving the general performance of the automatic AnEnPi classifier.

Although sequencing methods have been producing large amounts of data, the genome assemblies based on these data may be incomplete or approximate.⁵⁹ Low-quality assemblies result in low-quality annotations^{60,61} and promote both over- and underestimations of the numbers of genes.⁶² Considering the subject organism in a comparison, the incomplete assembly of its genome may lead to gene loss and thus generate a virtual specific gene in the query genome. It is possible to check the consistency of an event that is observed, but it is almost impossible to check the consistency of an event that is not observed. Thus, in this study, we did not take missed assemblies into account as a source of *false positives*. Only an in vitro analysis of our set of putative strictly specific enzymes can address this question.

For example, the existence of pseudogenes can be virtual or real, depending on whether they are the consequence of a mistake in the sequencing (induced frame shifts and premature stop codon) or assembling (gene cut) processes that lead to the inaccurate annotations.^{63–65} Indeed, many true pseudogenes have some form of biological activity, and thus their accurate annotation is potentially important to drug development.⁶⁵ We observed only 5 cases of a human pseudogene associated with the proteins of *L major*.

The homologies obtained by implementing HHpred and HHsearch have a higher confidence level compared with BLAST because HMMs consider the similarity in the amino acid sequence pairs from the *L major* query and the human sequence as well as the 2D and 3D structures of each sequence.

Enzyme essentiality

Enzymatic activity can play a key role in a metabolic pathway. Every pathway is associated with enzymes and their ECs. Therefore, a metabolic pathway is an oriented graph whose vertices are characterized by an enzyme and its associated chemical reaction.⁶⁶ A key enzyme is an enzyme that is contained in a path, lacks an alternative option, and plays an essential role in the survival of an organism. Consequently, the inhibition of a key enzyme necessarily results in the inhibition of the corresponding pathway and the organism's debilitation. However, if a pathway that does not play an important role in the survival of the organism is targeted, one cannot inhibit the pathway, and an organism can continue to reproduce normally. Thus, the key feature for exploring a metabolic pathway is to identify the enzymes that play a key role in a pathway and to determine whether they are essential for the survival of the parasite, ie, whether their inhibition is deleterious to that organism. In the context of drug development, the ideal drug should be the one that can inhibit the

activity of a specific enzyme in an essential pathway of the target organism, with no alternative route.

Functional specificity

Strict specificity, ie, the existence of a given enzyme in the parasite and not in the host, is obviously the best situation. However, functional specificity, where substrate specificity can be identified for the parasite form compared with the host form, can be contemplated as well. Actually, EC 1.3.1.71 (*L major*), which is associated with the steroid biosynthesis pathway, is an example of functionally specific enzyme that is homologous to the human enzyme EC 1.3.1.70. EC 1.3.1.71 could be explored for drug development as antifungal and antitrypanosomal agents.⁶⁷ Xu et al⁶⁸ showed that the knockdown of the gene encoding the 14- α -demethylase enzyme (EC 1.14.13.70), which is upstream of EC 1.3.1.71, in *L major* initially contributed to a dramatic change in the profile of the lipid composition in amastigotes and to virulence attenuation, which was reversed after a few weeks. Because the activity of EC 1.14.13.70 is common to the route of cholesterol and ergosterol syntheses in mammals and *Leishmania* spp, respectively, it is possible that *L major* got around the loss of EC 1.14.13.70 function in the knockdown parasites using sterols from the host. This escape strategy would be unlikely to occur if EC 1.3.1.71 was inhibited instead of the 14- α -demethylase enzyme activity in *Leishmania* spp, as EC 1.3.1.71 is inserted in a pathway of ergosterol synthesis that is specific to Trityps. We did not find other potentially reliable functionally specific cases in this study because additional alternative metabolite routes may be available to the parasite to sustain the function that was inhibited by drug treatment.⁶⁸ More recently, McCall et al⁶⁹ demonstrated the importance of 14- α -demethylase enzyme in *L donovani*.

Strict specificity

Strictly specific enzymes can be divided in targets with an a priori lower (the 32 non-key enzymes) and a higher (the 8 key enzymes) potential for drug development.

Considering the strictly specific enzymes with a higher potential for drug development, our analysis showed that 4 enzyme functions, namely, sterol 24-C-methyltransferase (EC 2.1.1.41), trypanothione synthetase (EC 6.3.1.9), pyruvate phosphate dikinase (EC 2.7.9.1), and mitochondrial RNA-editing ligase (EC 6.5.1.3), can be considered for the development of a drug cocktail.

Unlike humans who have cholesterol in their biological membranes, trypanosomatids use ergosterol. Therefore, trypanosomatids present distinct enzymes in their sterol biosynthesis pathway.⁷⁰ Sterol 24-C-methyltransferase (EC 2.1.1.41) catalyzes a methylation of carbon 24, which is fundamental for ergosterol biosynthesis. The 24-C-methyltransferase reaction is inhibited by 22,26-azasterol, which causes morphological changes and lysis to *Leishmania* spp.⁶⁷

Trypanothione synthetase (EC 6.3.1.9) is fundamental to the synthesis of trypanothione, a redox metabolite of trypanosomatids that is involved in a number of processes, such as the regulation of intracellular thiol redox balance, drug resistance, defense against chemical agents, and oxidative stress.⁷¹ The enzymes trypanothione synthetase and trypanothione reductase are involved in the trypanothione biosynthesis and metabolism pathways and are under study for the development of an alternative chemotherapy against trypanosomatids.⁷²

Pyruvate phosphate dikinase (EC 2.7.9.1) is an enzyme present in trypanosomatid glycosomes.⁷³ It was reported that this enzyme plays a key role in maintaining the balance of ATP vs adenosine diphosphate in the organelles in procyclic *T. brucei*⁷⁴ and in the gluconeogenesis process in the amastigotes of *L. mexicana*.⁷⁵ Therefore, it is believed that the pathway that includes pyruvate phosphate dikinase is central to the energetic metabolism of *L. major* and that this enzyme deserves to be considered for leishmanicidal drugs.

The L-complex has been isolated from *T. brucei* and *Leishmania tarentolae* mitochondria. At least 16 protein components have been identified, including REL1 and REL2. The precise function of REL1 and REL2, both annotated as EC 6.5.1.3, has been described by complementation of the knock-out strains.^{76,77} In *T. brucei*, the conditional disruption of REL1 is lethal in vivo as it affects both U-deletion and U-insertion editing resulting in an overall decrease in RNA size.⁷⁸ However, the loss of REL2 has no effect on viability or on editing.⁷⁹

These enzymes act in different metabolic pathways and cellular processes: sterol biosynthesis and membrane structure (EC 2.1.1.41), glutathione metabolism and oxidative stress response (EC 6.3.1.9), energy metabolism and carbon fixation pathways in prokaryote metabolism (EC 2.7.9.1), and RNA editing/posttranscriptional RNA processing (EC 6.5.1.3). Therefore, the simultaneous inhibition of these enzymes and consequently these pathways/cellular processes would likely promote an irreversible collapse of *L. major* cells.

Using a strategy based on target druggability, Crowther et al⁸⁰ identified the enzymes LmjF.27.1870 (EC 6.3.1.9), LmjF.36.2380, and LmjF.36.2390 (the 2 latter having EC 2.1.1.41) as being potential targets for drug development in agreement with our results. However, these enzymes were later disregarded because they were not found to be orthologous to the enzymes classified as essential in other model organisms (*Caenorhabditis elegans*, *Escherichia coli*, *M. tuberculosis*, and/or *Saccharomyces cerevisiae*) by bench experiments, perhaps these enzymes were not classified as essential because they did not exist in these organisms, as shown in Table 6 for *M. tuberculosis*.

By contrast, based on an interactome inference, Flórez et al⁸¹ identified 142 specific targets in the signaling network of *L. major* compared with *H. sapiens*. The only EC among our 8 strictly specific targets that we found in common with their list is specifically EC 2.1.1.41 (LmjF.36.2380 and LmjF.36.2390), which seems to contradict the statement by Crowther et al.⁸⁰

In addition, methyltransferase has been identified as a promising drug target in *Cryptococcus neoformans*.⁸² Moreover, Goto et al^{83,84} described a vaccine candidate for the enzyme sterol 24-C-methyltransferase against visceral leishmaniasis that is effective against *L. infantum* and *L. donovani*.

Unlike Crowther et al,⁸⁰ we evaluated the essentiality of an enzyme by assessing its centrality in a metabolic pathway in the specific case of *L. major* using its metabolic representation in KEGG.

Thus, the 8 strictly specific enzymes represented by the 4 ECs that we found to be essential are also reported as essential in the literature. By contrast, looking into Research and Training in Tropical Diseases (TDR; <http://tdrtargets.org/>), we found that 5 of the enzymes are considered essential for one or more trypanosomatids, 2 (LmjF.36.2380 and LmjF.36.2390) are not considered essential, and 1 (LmjF.36.4300) does not have data on whether it is essentially associated.⁸⁵

Gene redundancy and parasite resistance to drugs

We verified that, with the exception of EC 2.7.9.1, which has only 1 associated enzyme sequence, the other 3 EC numbers for strictly specific targets (EC 2.1.1.41, EC 6.5.1.3, and EC 6.3.1.9) have more than 1 gene representative in the *L. major* genome. Gene redundancy for a same enzymatic function (see Table S2) is a potential source of drug resistance because a given inhibitor may inhibit a given enzyme, but not its putative isoforms, which could also be present.⁸⁶ Drug resistance may involve several different mechanisms, and a common solution to this problem is to increase the number of enzymes that are targeted by the inhibitors at the same time.

The process of gene accumulation, which has been called *pyramidation*,⁸⁷ is routinely used in classical breeding for plant resistance where it was first implemented.^{88,89} *Pyramidation* lowers the likelihood of virulence adaptation by a parasite because the corresponding accumulation of virulence genes becomes unsustainable in the given environmental conditions.

The idea of accumulating a number of targets that are simultaneously inhibited for a better disease control is a variation of the *gene-for-gene* relationship that has been described by Flor⁹⁰ and can be referred to as the *gene-for-inhibitor* concept.⁹¹ Thus, formulating drugs into a cocktail should overcome parasite resistance; this practice is actually⁹² already in use, as seen in the "Introduction" section. However, the formulation of a drug combination should also account for the dose-limiting negative side effects to normal cells and should protect the integrity of the host immune system. Thus, we sought to address this question here by identifying only the enzyme targets that are specific to *L. major*, as targeting these candidates would most likely minimize the deleterious side effects from a therapeutic combination to patients. Of course, the results outlined in this report must still be confirmed by in vitro experimentation. Another question to be assessed in the future is the

requirement for safety studies for each of the constituents of the combination treatment, independently and in combination, prior to the successful registration of that combination treatment; however, this question is beyond the scope of this report, which is mainly exploratory according to our current knowledge.

Drugs for key strictly specific L major enzymes

Drug reposition is a preferred route in drug development due to its savings in time and money. Unfortunately, an examination of TDR Targets did not show any approved drugs for the enzyme targets described here, but some inhibitors appear to be available at least for the experimental stages in *Leishmania* or other biological species. A search in the literature uncovered that (1) azasterols are inhibitors of sterol 24-C-methyltransferase (EC 2.1.1.41) in *Leishmania*^{67,92} and (2) azo dye naphthalene-like compounds, similar to suramin, are inhibitors of REL (EC 6.5.1.3).⁹³ Two of these compounds were later tested on the whole editosome and showed strong inhibition of U-deletion in RNA editing.⁹⁴ GW5074, mitoxantrone, NF 023, protoporphyrin IX, and D-sphingosine were also identified as REL inhibitors, with IC₅₀ values ranging from 1 to 3 μM.⁹⁵ (3) Phosphonopyruvate (ID: DB02522 in DrugBank) has been characterized as an inhibitor⁹⁶ of pyruvate phosphate dikinase (EC 2.7.9.1) as well as flavones analogs. It has also been reported that flavone analogs act by specifically targeting the ATP-binding site of the enzyme in a fungal system.⁹⁷ Several phenolic compounds were also proposed for the inhibition of pyruvate phosphate dikinase in *Wolbachia*.⁹⁸ (4) A bis-benzyl diamine derivative, as well as MOL2008, was proposed as new molecular scaffolds for drug development, which proved to be active at EC₅₀ = 12 μM on *L. infantum* promastigotes; these compounds depleted intracellular trypanothione by inhibiting trypanothione synthetase (EC 6.3.1.9) in treated parasites, confirming the on-target activity of these compounds.⁹⁹

Under the hypothesis that the susceptibility of *Leishmania* would be stable in vitro to a combination of the compounds listed above, it would be possible to optimize their scaffold structure to reduce the overall cocktail toxicity to a human cell model. The engineering of this process is now facilitated by in silico modeling. A search for 3D models in PDB with sequences that are homologous to LmjF.36.2380, LmjF.36.2390, LmjF.11.1000, LmjF.23.0460, LmjF.27.1870, LmjF.36.4300, LmjF.01.0590, and LmjF.20.1730 showed that only the sequences of EC 2.7.9.1 and EC 6.3.1.9 were suitable for 3D modeling because their identity scores were larger than 39%, reaching up to a maximum of 97%. However, in the case of EC 6.5.1.3 and EC 2.1.1.41, the identity scores were too low (<25%) and the number of gaps were too high for successful 3D modeling.

Distribution of key strictly specific L major enzymes among human parasites

Sterol 24-C-methyltransferase (EC 2.1.1.41) is a component of the ergosterol pathway. This pathway has been extensively studied for drug development against fungi, and ergosterol pathway inactivation has been demonstrated as efficient and stable. The fact that sterol 24-C-methyltransferase can be shared with fungi means that drug reposition from fungi to Trypanosomatidae may occur, as was the case with fluconazole. Pyruvate phosphate dikinase (EC 2.7.9.1) is highly specific to Trypanosomatidae, and EC 2.7.9.1 was only found to be shared with *M. tuberculosis*; however, the sequences did not align, which suggests that the enzymes of both lineages are analogous. The sequences of trypanothione synthetase (EC 6.3.1.9) of *T. cruzi* showed identities between 37% and 58% with those of *L. major*. These data agree with the literature, which demonstrates the expression of this enzyme and the presence of the corresponding trypanothione metabolism in *T. cruzi*.⁸⁰ As expected, *L. major*, which belongs to Trityps, had a greater phylogenetic affinity with *T. cruzi* and *T. brucei* (Euglenozoa) than with *P. falciparum*, which belongs to a different phylum (Apicomplexa). The status of sequence analogy that we gave to EC 6.5.1.3 (REL1) of *P. falciparum* and *T. gondii* is not surprising, considering that both species belong to the clade of Alveolates (Chromalveolates), whereas Trypanosomatidae belongs to Euglenozoa (Excavates), which are as far away from one another as they are from Animalia (Unikonts) in evolutionary terms (<http://tolweb.org/Eukaryotes/3>). However, the status of analogy only depends on the EC annotation. In that respect, one could ask whether an RNA-editing ligase that is involved in trans-splicing in Trityps should have the same EC (EC 6.5.1.3) as a tRNA-splicing ligase that is involved in cis-splicing and tRNA repair, as the biological consequences of the 2 reactions are different in both cases. Interestingly, REL1 is conserved across the subclass of Metakinetoplastina, which shows its ancient emergence within Excavates. Unfortunately, we did not find a complete genome sequence for the *Diplonema* or *Euglena* from higher taxa to check the precise level of life history in which REL emerged. Apart from parasites, Metakinetoplastina also includes nonparasitic free-living forms, such as *B. saltans*, which shows that REL function predates parasitic adaptation. Actually, REL is encoded by an *informational gene* and belongs to the primeval functions that were used by life as soon as the RNA world came into existence,¹⁰⁰ and it has been shown to have been invented at least 2 times independently.

Expression pattern of key strictly specific L major enzymes

Interestingly, the expression of LmjF.01.0590, LmjF.11.1000, LmjF.20.1730, LmjF.23.0460, LmjF.27.1870, LmjF.36.2380, LmjF.36.2390, and LmjF.36.4300 is described as being

constitutive in the promastigote (blood) and amastigote (intracellular) forms of *L. major*,^{101–103} which makes these sequences suitable targets for drugs that are effective against both parasite forms.

Conclusions

By applying a comparative method of enzyme function for the pathogenic relationship between *L. major* and humans, we succeeded in restricting the number of specific enzyme targets that we believe are suitable for drug development. Of the 67 sequences (accounting for 42 enzymatic activities) that were classified as specific to *L. major* compared with *H. sapiens* by AnEnPi, only 40 (23 ECs) were *strictly* specific, as 27 enzymes (19 ECs) were disregarded for being ambiguous, *functionally* specific, or analogous. Among the list of 40 strictly specific enzymes, another 32 sequences (19 ECs) were also disregarded for not being *essential* in *L. major* metabolism. Finally, we uncovered 4 enzymes (sterol 24-C-methyltransferase, RNA-editing ligase, pyruvate phosphate dikinase, and trypanothione synthetase) that are suitable for drug development with the purpose of minimizing treatment toxicity to the host. The method is systematic, includes a process for determining *false positives*, can be automated, and can be applied to the investigation of other host-parasite relationships.

Acknowledgements

The authors thank members of the *Plataforma de Bioinformática Fiocruz RPT04-A/RJ* for technological support and David William Provance Jr for reading and editing the manuscript.

Author Contributions

LCC, NC, and ACRG conceived and designed the experiments; wrote the first draft of the manuscript. LCC, NC, CRL, MA-F, and ACRG analyzed the data. LCC and NC jointly developed the structure and arguments of the paper and wrote it with the contribution of MA-F. LCC, NC, CRL, ACRG, MA-F, PT, and PD agree with manuscript results and conclusions. LCC, NC, AFF, MA-F, PT, and PD made critical revisions and approved final version. All authors reviewed and approved the final manuscript.

REFERENCES

- Brazil RP, Rodrigues AAF, Filho JDA. Sand fly vectors of *Leishmania* in the Americas—a mini review. *Entomol Ornithol Herpetol Curr Res*. 2015;4:2.
- Georgiadou SP, Makaritsis KP, Dalekos GN. Leishmaniasis revisited: current aspects on epidemiology, diagnosis and treatment. *J Transl Int Med*. 2015;3:43–50.
- World Health Organization. Leishmaniasis fact sheet no. 375. <http://www.who.int/mediacentre/factsheets/fs375/en/>. Published 2016. Accessed December 10, 2016.
- Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2095–2128.
- Sundar S, Chakravarty J. An update on pharmacotherapy for leishmaniasis. *Expert Opin Pharmacother*. 2015;16:237–252.
- Moën MD, Lyseng-Williamson KA, Scott LJ. Liposomal amphotericin B: a review of its use as empirical therapy in febrile neutropenia and in the treatment of invasive fungal infections. *Drugs*. 2009;69:361–392.
- Wyllie S, Cunningham ML, Fairlamb AH. Dual action of antimonial drugs on thiol redox metabolism in the human pathogen *Leishmania donovani*. *J Biol Chem*. 2004;279:39925–39993.
- Baiocco P, Colotti G, Franceschini S, Ilari A. Molecular basis of antimony treatment in leishmaniasis. *J Med Chem*. 2009;52:2603–2612.
- Frézard F, Demicheli C, Ribeiro RR. Pentavalent antimonials: new perspectives for old drugs. *Molecules*. 2009;14:2317–2336.
- Chawla B, Jhingran A, Panigrahi A, Stuart KD, Madhubala R. Paromomycin affects translation and vesicle-mediated trafficking as revealed by proteomics of paromomycin -susceptible -resistant *Leishmania donovani*. *PLoS ONE*. 2011;6:e26660.
- Rakotomanga M, Blanc S, Gaudin K, Chaminade P, Loiseau PM. Miltefosine affects lipid metabolism in *Leishmania donovani* promastigotes. *Antimicrob Agents Chemother*. 2007;51:1425–1430.
- Verma NK, Dey CS. Possible mechanism of miltefosine-mediated death of *Leishmania donovani*. *Antimicrob Agents Chemother*. 2004;48:3010–3015.
- World Health Organization. http://apps.who.int/iris/bitstream/10665/44412/1/WHO_TRS_949_eng.pdf. Published 2010. Accessed February 10, 2017.
- Warrilow AG, Parker JE, Kelly DE, Kelly SL. Azole affinity of sterol 14 α -demethylase (CYP51) enzymes from *Candida albicans* and *Homo sapiens*. *Antimicrob Agents Chemother*. 2013;57:1352–1360.
- Palumbo E. Treatment strategies for mucocutaneous leishmaniasis. *J Glob Infect Dis*. 2010;2:147–150.
- Sousa AQ, Frutuoso MS, Moraes EA, Pearson RD, Pompeu MM. High-dose oral fluconazole therapy effective for cutaneous leishmaniasis due to *Leishmania (Vianna) braziliensis*. *Clin Infect Dis*. 2011;53:693–695.
- Kandpal M, Tekwani BL, Chauhan PMS, Bhaduri AP. Correlation between inhibition of growth and arginine transport of *Leishmania donovani* promastigotes in vitro by diamidines. *Life Sci*. 1996;59:75–80.
- Reguera R, Balana Fouce R, Cubria JC, Alvarez Bujidos ML. Putrescine uptake inhibition by aromatic diamidines in *Leishmania infantum* promastigotes. *Biochem Pharmacol*. 1994;47:1859–1866.
- Basselin M, Coombs GH, Barrett MP. Putrescine and spermidine transport in *Leishmania*. *Mol Biochem Parasitol*. 2000;109:37–46.
- Ouellette M, Drummel-Smith J, Papadopoulou B. Leishmaniasis: drugs in the clinic, resistance and new developments. *Drug Resist Updat*. 2004;7:257–266.
- Khan I, Khan M, Umar MN, Oh D-H. Attenuation and production of the amphotericin B-resistant *Leishmania tropica* strain. *Jundishapur J Microbiol*. 2016;9:e32159.
- Arrowsmith J. A decade of change. *Nat Rev Drug Discov*. 2012;11:17–18.
- Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9:203–214.
- Kinch MS, Haynesworth A, Kinch SL, Hoyer D. An overview of FDA-approved new molecular entities: 1827-2013. *Drug Discov Today*. 2014;19:1033–1039.
- Bhopal A, Callender T, Knox AF, Regmi S. Strength in numbers? Grouping, fund allocation and coordination amongst the neglected tropical diseases. *J Glob Health*. 2013;3:020302.
- Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod*. 2012;75:311–335.
- Bombelles T, Coaker H. Neglected tropical disease research: rethinking the drug discovery model. *Future Med Chem*. 2015;7:693–700.
- Bull SC, Doig AJ. Properties of protein drug target classes. *PLoS ONE*. 2015;10:e0117955.
- Capriles PVSZ, Baptista LPR, Guedes IA, et al. Structural modeling and docking studies of ribose 5-phosphate isomerase from *Leishmania major* and *Homo sapiens*: a comparative analysis for Leishmaniasis treatment. *J Mol Graph Model*. 2015;55:134–147.
- Gabaldón T, Koonin EV. Functional and evolutionary implications of gene orthology. *Nat Rev Genet*. 2013;14:360–366.
- Fitch WM. Distinguishing homologous from analogous proteins. *Syst Zool*. 1970;19:99–113.
- Otto TD, Guimarães ACR, Degraeve WM, de Miranda AB. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics*. 2008;9:544.
- Galperin MY, Walker DR, Koonin EV. Analogous enzymes: independent inventions in enzyme evolution. *Genome Res*. 1998;8:779–790.
- Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. 2006;34:D354–D357.
- Guimarães AC, Otto TD, Alves-Ferreira M, Miranda AB, Degraeve WM. In silico reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*. *Genet Mol Res*. 2008;7:872–882.
- Alves-Ferreira M, Guimarães ACR, Capriles PVSZ, Dardenne LE, Degraeve WM. A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information. *Mem Inst Oswaldo Cruz*. 2009;104:1100–1110.

37. Gomes MR, Guimarães ACR, de Miranda AB. Specific and nonhomologous isofunctional enzymes of the genetic information processing pathways as potential therapeutic targets for tritryps. *Enzyme Res.* 2011;2011:543912.
38. Flicek P, Amode MR, Barrell D, et al. Ensembl 2014. *Nucleic Acids Res.* 2014;42:D749–D755.
39. Aslett M, Aurrecochea C, Berriman M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 2009;38:D457–D462.
40. Pruitt KD, Harrow J, Harte RA, et al. The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* 2009;19:1316–1323.
41. Farrell CM, O'Leary NA, Harte RA, et al. Current status and new features of the consensus coding sequence database. *Nucleic Acids Res.* 2014;42:D865–D872.
42. The UniProt Consortium. Activities at the universal protein resource (UniProt). *Nucleic Acids Res.* 2014;42:D191–D198.
43. Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42:222–230.
44. Marchler-Bauer A, Anderson JB, Derbyshire MK, et al. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 2007;35:D237–D240.
45. Simossis VA, Heringa J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* 2005;33:W289–W294.
46. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005;21:951–960.
47. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 2005;33:W244–W248.
48. Neron B, Menager H, Maufrais C, et al. Moby: a new full web bioinformatics framework. *Bioinformatics.* 2009;25:3005–3011.
49. Rimmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods.* 2011;9:173–175.
50. Lima CR, Carels N, Guimaraes ACR, Tufféry P, Derreumaux P. In silico structural characterization of protein targets for drug development against *Trypanosoma cruzi*. *J Mol Model.* 2016;22:244.
51. Rost B. Twilight zone of protein sequence alignments. *Protein Eng.* 1999;12:85–94.
52. Panigrahi AK, Schnauffer A, Ernst NL, et al. Identification of novel components of *Trypanosoma brucei* editosomes. *RNA.* 2003;9:484–492.
53. Panigrahi AK, Gygi SP, Ernst NL, et al. Association of two novel proteins, TbMP52 and TbMP48, with the *Trypanosoma brucei* RNA editing complex. *Mol Cell Biol.* 2001;21:380–389.
54. Mehta V, Sen R, Moshiri H, Salavati R. Mutational analysis of *Trypanosoma brucei* RNA editing ligase reveals regions critical for interaction with KREPA2. *PLoS ONE.* 2015;10:e0120844.
55. Simpson AG, Stevens JR, Lukes J. The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol.* 2006;22:168–174.
56. McWilliam H, Li W, Uludag M, et al. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* 2013;41:W597–W600.
57. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–1874.
58. Wilming LG, Gilbert JGR, Howe K, Trevanion S, Hubbard T, Harrow JL. The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.* 2008;36:D753–D760.
59. Denton JF, Lugo-Martinez J, Tucker AE, Schridder DR, Warren WC, Hahn MW. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput Biol.* 2014;10:e1003998.
60. Phillippy AM, Schatz MC, Pop M. Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol.* 2008;9:R55.
61. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. *Nat Methods.* 2011;8:61–65.
62. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* 2013;30:1987–1997.
63. Echols N, Harrison P, Balasubramanian S, et al. Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes. *Nucleic Acids Res.* 2002;30:2515–2523.
64. Balakirev ES, Ayala FJ. Pseudogenes: are they "junk" or functional DNA? *Annu Rev Genet.* 2003;37:123–151.
65. Pei B, Sisu C, Frankish A, et al. The GENCODE pseudogene resource. *Genome Biol.* 2012;13:R51.
66. Selkov E, Grechkin Y, Mikhailova N, Selkov E. MPW: the metabolic pathways database. *Nucleic Acids Res.* 1998;26:43–45.
67. de Souza W, Rodrigues JCF. Sterol biosynthesis pathway as target for anti-trypanosomatid drugs. *Interdiscip Perspect Infect Dis.* 2009;6:42502.
68. Xu W, Hsu F-F, Baykal E, Huang J, Zhang K. Sterol biosynthesis is required for heat resistance but not extracellular survival in *Leishmania*. *PLoS Pathog.* 2014;10:e1004427.
69. McCall L-I, El Aroussi A, Choi JY, et al. Targeting ergosterol biosynthesis in *Leishmania donovani*: essentiality of sterol 14 α -demethylase. *PLoS Negl Trop Dis.* 2015;9:e0003588.
70. Roberts CW, McLeod R, Rice DW, Ginger M, Chance ML, Goad LJ. Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Mol Biochem Parasitol.* 2003;126:129–142.
71. Fairlamb AH, Blackburn P, Ulrich P, Chait BT, Cerami A. Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. *Science.* 1985;227:1485–1487.
72. Leroux AE, Krauth-Siegel RL. Thiol redox biology of trypanosomatids and potential targets for chemotherapy. *Mol Biochem Parasitol.* 2016;206:67–74.
73. Bringaud F, Baltz D, Baltz T. Functional and molecular characterization of a glycosomal PPI-dependent enzyme in trypanosomatids: pyruvate, phosphate dikinase. *Proc Natl Acad Sci U S A.* 1998;95:7963–7968.
74. Deramchia K, Morand P, Biran M, et al. Contribution of pyruvate phosphate dikinase in the maintenance of the glycosomal ATP/ADP balance in the *Trypanosoma brucei* procyclic form. *J Biol Chem.* 2014;289:17365–17378.
75. Rodriguez-Contreras D, Aslan H, Feng X, et al. Regulation and biological function of a flagellar glucose transporter in *Leishmania mexicana*: a potential glucose sensor. *FASEB J.* 2015;29:11–24.
76. Gao G, Simpson AM, Kang X, et al. Functional complementation of *Trypanosoma brucei* RNA in vitro editing with recombinant RNA ligase. *Proc Natl Acad Sci.* 2005;102:4712–4717.
77. Kang X, Gao G, Rogers K, Falick AM, Zhou S, Simpson L. Reconstitution of full-round uridine-deletion RNA editing with three recombinant proteins. *Proc Natl Acad Sci.* 2006;103:13944–13949.
78. Schnauffer A, Panigrahi AK, Panicucci B, et al. An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science.* 2001;291:2159–2162.
79. Gao G, Simpson L. Is the *Trypanosoma brucei* REL1 RNA ligase specific for U-deletion RNA editing, and is the REL2 RNA ligase specific for U-insertion editing? *J Biol Chem.* 2003;278:27570–27574.
80. Crowther GJ, Shanmugam D, Carmona SJ, et al. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. *PLoS Negl Trop Dis.* 2010;4:e804.
81. Flórez AF, Park D, Bhak J, et al. Protein network prediction and topological analysis in *Leishmania major* as a tool for drug target selection. *BMC Bioinformatics.* 2010;11:484.
82. Nes WD, Zhou W, Ganapathy K, et al. Sterol 24-C-methyltransferase: an enzymatic target for the disruption of ergosterol biosynthesis and homeostasis in *Cryptococcus neoformans*. *Arch Biochem Biophys.* 2009;481:210–218.
83. Goto Y, Bogatzki LY, Bertholet S, Coler RN, Reed SG. Protective immunization against visceral leishmaniasis using *Leishmania* sterol 24-C-methyltransferase formulated in adjuvant. *Vaccine.* 2007;25:7450–7458.
84. Goto Y, Bhatia A, Raman VS, et al. *Leishmania infantum* sterol 24-c-methyltransferase formulated with MPL-SE induces cross-protection against *L. major* infection. *Vaccine.* 2009;27:2884–2890.
85. Agüero F, Al-Lazikani B, Aslett M, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov.* 2008;7:900–907.
86. Chacko AD, McDade SS, Chanduly S, et al. Expression of the SEPT9_i4 isoform confers resistance to microtubule-interacting drugs. *Cell Oncol.* 2012;35:85–93.
87. Bury M, Andolfi A, Rogister B, et al. Fusicoocin A, a phytotoxic carbocyclic diterpene glucoside of fungal origin, reduces proliferation and invasion of glioblastoma cells by targeting multiple tyrosine kinases. *Transl Oncol.* 2013;6:112–123.
88. Rispaill N, Dita M-A, González-Verdejo C, et al. Plant resistance to parasitic plants: molecular approaches to an old foe. *New Phytol.* 2007;173:703–712.
89. Keane PJ. Horizontal or generalized resistance to pathogens in plants. In: Cumagun, CJR ed. *Plant Pathology*. Rijeka, Croatia: InTech; 2012:327–362.
90. Flor HH. Current status of the gene-fob-gene concept. *Annu Rev Phytopathol.* 1971;9:275–296.
91. Arteaga CL, Engelman JA. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell.* 2014;25:282–303.
92. Magaraci F, Jimenez Rodrigues C, Rodrigues JCF, et al. Azasterols as inhibitors of sterol 24-methyltransferase in *Leishmania* species and *Trypanosoma cruzi*. *J Med Chem.* 2003;46:4714–4727.
93. Amaro RE, Schnauffer A, Interthal H, Hol W, Stuart KD, McCammon JA. Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proc Natl Acad Sci.* 2008;105:17278–17283.
94. Moshiri H, Acoca S, Kala S, et al. Naphthalene-based RNA editing inhibitor blocks RNA editing activities and editosome assembly in *Trypanosoma brucei*. *J Biol Chem.* 2011;286:14178–14189.

95. Liang S, Connell GJ. Identification of specific inhibitors for a trypanosomatid RNA editing reaction. *RNA*. 2010;16:2435–2441.
96. Herzberg O, Chen CCH, Liu S, et al. Pyruvate site of pyruvate phosphate dikinase: crystal structure of the enzyme-phosphonopyruvate complex, and mutant analysis. *Biochemistry*. 2002;41:780–787.
97. Wu C, Dunaway-Mariano D, Mariano PS. Design, synthesis, and evaluation of inhibitors of pyruvate phosphate dikinase. *J Org Chem*. 2013;78:1910–1922.
98. Palayam M, Lakshminarayanan K, Radhakrishnan M, Krishnaswamy G. Preliminary analysis to target pyruvate phosphate dikinase from wolbachia endosymbiont of *Brugia malayi* for designing anti-filarial agents. *Interdiscip Sci Comput Life Sci*. 2012;4:74–82.
99. Benítez D, Medeiros A, Fiestas L, et al. Identification of novel chemical scaffolds inhibiting trypanothione synthetase from pathogenic trypanosomatids. *PLoS Negl Trop Dis*. 2016;10:e0004617.
100. Kim KM, Caetano-Anollés G. The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol Biol*. 2011;11:140.
101. Oza SL, Shaw MP, Wyllie S, et al. Trypanothione biosynthesis in *Leishmania major*. *Mol Biochem Parasitol*. 2005;139:107–116.
102. Leifso K, Cohen-Freue G, Dogra N, et al. Genomic and proteomic expression analysis of *Leishmania* promastigote and amastigote life stages: the *Leishmania* genome is constitutively expressed. *Mol Biochem Parasitol*. 2007;152:35–46.
103. Opperdoes FR, Coombs GH. Metabolism of *Leishmania*: proven and predicted. *Trends Parasitol*. 2007;23:149–158.



Research paper

Specific enzyme functionalities of *Fusarium oxysporum* compared to host plants



Larissa Catharina, Nicolas Carels*

Laboratório de Modelagem de Sistemas Biológicos, Instituto Nacional de Ciência e Tecnologia de Inovação em Doenças de Populações Negligenciadas (INCT-IDPN), Centro de Desenvolvimento Tecnológico em Saúde (CDTS), Fundação Oswaldo Cruz (Fiocruz), Av. Brasil, 4036, Prédio da Expansão, 8º andar, sala 814, CEP: 21040-361 Rio de Janeiro, Brazil

ARTICLE INFO

Keywords:

Fusariosis
Arabidopsis
Oilseed plants
Specific enzymes
Molecular targets

ABSTRACT

The genus *Fusarium* contains some of the most studied and important species of plant pathogens that economically affect world agriculture and horticulture. *Fusarium* spp. are ubiquitous fungi widely distributed in soil, plants as well as in different organic substrates and are also considered as opportunistic human pathogens. The identification of specific enzymes essential to the metabolism of these fungi is expected to provide molecular targets to control the diseases they induce to their hosts. Through applications of traditional techniques of sequence homology comparison by similarity search and Markov modeling, this report describes the characterization of enzymatic functionalities associated to protein targets that could be considered for the control of root rots induced by *Fusarium oxysporum*. From the analysis of 318 *F. graminearum* enzymes, we retrieved 30 enzymes that are specific of *F. oxysporum* compared to 15 species of host plants. By comparing these 30 specific enzymes of *F. oxysporum* with the genome of *Arabidopsis thaliana*, *Brassica rapa*, *Glycine max*, *Jatropha curcas* and *Ricinus communis*, we found 7 key specific enzymes whose inhibition is expected to affect significantly the development of the fungus and 5 specific enzymes that were considered here to be secondary because they are inserted in pathways with alternative routes.

1. Introduction

The genus *Fusarium* is known to include pathogenic fungi with a wide range of hosts and infection strategies (Goswami and Kistler, 2004) whose endemic areas are mostly located in tropical and subtropical countries (O'Donnell et al., 2008). The pathogenic fungus *Fusarium oxysporum* is known to cause root rot in > 100 plant species (Agris, 2005), including *Arabidopsis thaliana*, *Jatropha curcas*, *Glycine max*, *Brassica rapa* and *Ricinus communis* (Enya et al., 2008; Michielse and Rep, 2009; Kumar and Kaushik, 2013; Trusov et al., 2013; Anjani et al., 2014); the last four being important oilseeds for biodiesel production.

After germination and penetration into host roots, *F. oxysporum* enters the vascular system and uses xylem vessels to quickly colonize its plant host and promote vessel clogging. The successful colonization of

host plants by *F. oxysporum* leads to necrosis of infected tissues, subsequent vascular vessel collapse and plant death (Beckman, 1987). This process requires a number of tightly controlled biological steps: (i) recognition of roots through unknown host signals; (ii) fixation on root surface by hyphae and differentiation into penetrating structures; (iii) penetration into root cortex and degradation of host physical barriers; (iv) adaptation to the hostile environment created by infected plant tissues; (v) proliferation of hyphae within xylem vessels and secretion of virulence determinants and (vi) production of microconidia (Pietro et al., 2003).

Because of host-parasite coevolution, *formae speciales* were recognized into *F. oxysporum* based on symptoms, epidemiology and susceptibility of hosts (Armstrong and Armstrong, 1981; Nelson et al., 1994). > 150 different *formae speciales* were identified from isolates of *F. oxysporum* according to its host range (Baayen et al., 2000; Michielse

Abbreviations: Ap4A, diadenosine tetraphosphate; CAT, carnitine acetyltransferase; CDP-choline, cytidine diphosphate choline; CWDEs, cell walls degrading enzymes; DcpS, decapping scavenger enzyme; EC, enzyme commission numbers; HCCS, holocytochrome-c synthase; HGT, horizontal gene transfer; HHblits, HMM-based lightning-fast iterative sequence search; HMM, hidden Markov model; IC50, half maximal inhibitory concentration; IDO, indoleamine 2,3-dioxygenase; KEGG, Kyoto Encyclopedia of Genes and Genomes database; L-Met, L-methionine; L-Trp, L-tryptophan; NAD, nicotinamide adenine dinucleotide; N-SMase, neutral sphingomyelinases; *nr*, non-redundant section of GenBank; SMA, Spinal Muscular Atrophy; UDP-GlcNAc, UDP-N-acetylglucosamine

* Corresponding author.

E-mail addresses: larissa.costa@cmts.fiocruz.br (L. Catharina), nicolas.carels@cmts.fiocruz.br (N. Carels).

<https://doi.org/10.1016/j.gene.2018.07.003>

Received 18 September 2017; Received in revised form 14 May 2018; Accepted 1 July 2018

Available online 05 July 2018

0378-1119/ © 2018 Published by Elsevier B.V.

and Rep, 2009). In addition, *formae speciales* may also be subdivided into physiological races according to their virulence.

Fungi of the genus *Fusarium* also produce a range of toxic secondary metabolites (mycotoxins), such as (i) trichothecenes that inhibit protein translation and have eliciting activities that stimulate plant defenses and promote plant cell death. For instance, agricultural products contaminated by fumonisins are unsuitable for human or animal consumption (Desmond et al., 2008; Ma et al., 2013; Woloshuk and Shim, 2013).

In addition, *Fusarium* spp. secrete many enzymes that may contribute to virulence, such as those that act on physical and chemical barriers induced by defense responses (de Sain and Rep, 2015) and can damage them by cell walls degrading enzymes (CWDEs). Upon contact with the wall of host cells, *Fusarium* spp. synthesize extracellular cell wall degradation enzymes following a defined sequence of molecular steps (Jones et al., 1972), whose genes are up-regulated by the presence of enzyme substrate and down-regulated by processes of catabolic repression (Cooper and Wood, 1973).

As the fungus attempts to invade its host, it may face toxic secondary metabolites with antimicrobial activity whose synthesis is induced by plant defense mechanisms (Morrissey and Osbourn, 1999). Successful colonization of host plants by *F. oxysporum* leads to necrosis of infected tissues, the subsequent collapse of vascular vessels and plant death. The genetics of plant resistance to *F. oxysporum* is complex and difficult to transfer by hybridization, therefore, there is interest to uncover the molecular basis of the plant's innate immunity against these pathogens (Berrocal-Lobo and Molina, 2008).

Chitinase is a predominant enzyme in the biodegradation process of chitin, which in turn is an essential component of the fungal cell wall (Viterbo et al., 2002). By contrast, chitin is not a component of plant cell wall, thus plant chitinase genes are key genes for plant defense and can be activated in response to biotic stresses (Cheng et al., 2017).

The application of fungicides is an important factor of crop productivity. Fungicides can be applied directly to the field or during storage and transport. Unfortunately, the phenomenon of microorganisms resistance to the chemicals used in agriculture is constantly observed (Wolny-Kołodka, 2014). In ornamental plants, for instance, the number of registered chemicals has declined over the years. The resistance of *F. oxysporum* to benzimidazole makes difficult to control bulb rot and wilt in several susceptible plant cultivars (Gullino and Garibaldi, 2005). In addition, the fungicidal activity may also affect other organisms that contribute positively to the crop.

In this context, the identification of alternative strategies, such as the inactivation of enzymes specific to the fungal parasite participating in its essential metabolic pathways, is indispensable in the attempt to propose new inhibitors with increased selective activity.

2. Methodology

2.1. Background

The Kyoto Encyclopedia of Genes and Genomes database (KEGG, <http://www.genome.jp/kegg/>) was used to retrieve enzymatic annotations and their corresponding protein sequences (version 74.0, April 1, 2015) in the organisms considered below. Enzyme reactions were inferred to be specific of *Fusarium graminearum* compared to *A. thaliana* when enzyme commission numbers (EC) from the KEGG list of *F. graminearum* were not found in that of *A. thaliana*. Thus, the method of specificity diagnostic is merely a comparison of EC lists. However, the aspect that is less simple is how to curate these lists to eliminate the false positives of specific enzymes. False positives can result from incomplete EC annotation and enzymes that are classified as not present in one organism when they are actually present. The question, I how to proceed to verify that enzymes present in a parasitic organism are effectively not present in the host of this parasite. This operation can be done at the level of protein themselves as well as at the level of the host

genome. At the level of proteins, an improvement is to complement the detection of sequence similarity (BLAST) by Markov chain comparison, which is able to detect remote homologies by comparing a given sequence from the parasite with Markovian models stored in database that take primary sequence and profile of secondary structures into consideration. At the level of host genome, the question is to confirm that its protein annotation did not skip any genome sequence that could encode a homologous protein of the parasite sequence under consideration; this is typically done by comparison of similarity among sequences (BLAST). When putatively true positives are diagnosed, it is interesting to tentatively report on their potentiality as target for inhibition by looking at their centrality in pathway maps.

In nature, there is a *forma speciale* of *F. oxysporum* that is pathogenic to *A. thaliana*, however, neither the genome of *F. oxysporum* nor its EC annotation were available in KEGG when the study was performed. Therefore, EC data of the fungus *F. graminearum* were used instead since its genomic data were available in KEGG. This assessment was based on the observation by Ma et al. (2010) that core orthologous genes of *F. graminearum* and *F. oxysporum* are 85% similar, on the average. The two lists of ECs of *F. graminearum* and *A. thaliana* were taken from KEGG and compared to find the subset of ECs that are specific to *F. graminearum*. Then, we used (i) the KEGG annotations to retrieve protein sequences from the accession names associated to specific ECs from *F. graminearum* and (ii) the specific sequences from *F. graminearum* to retrieve their counterparts from *F. oxysporum* (see Section 2.3 below). Subsequently, *F. oxysporum* enzymes counterparts of *F. graminearum* were confirmed for their specificity by sequence comparison to *A. thaliana* using BLASTp. Thus, the rationale to find specific EC's from *F. oxysporum* was to use the list of *F. graminearum* EC from KEGG, compare it to the list to of Arabidopsis' ECs, find the ECs specific to *F. graminearum*, eliminate false positive and transfer the annotation of specific enzymes from *F. graminearum* to *F. oxysporum* using BLASTp.

2.2. Search for orthologous enzymes in *F. oxysporum*

With the UniProtKB equivalent of the ECs considered to be specific to *F. graminearum*, we retrieved their orthologs in *F. oxysporum* (strain Fo5176) using OrthoDB (Zdobnov et al., 2016; <http://www.orthodb.org/>). When an enzymatic function was lacking, it was transferred from *F. graminearum* to *F. oxysporum* (strain Fo5176) under the guideline of the following thresholds: e-value $\leq 10^{-4}$, identity $\geq 35\%$ and coverage between similarity region and query sequence $\geq 75\%$ (Lima et al., 2016).

2.3. Hidden Markov model

BLAST is a high throughput tool of sequence homology comparison, however, sequences can still be homologous at the level of the tertiary structure (3D) without recognizable homology at the level of the primary one (sequence). The twilight zone is the homology level (30–35%) at which the homology classification with BLAST becomes ambiguous (Rost, 1999). The hidden Markov model (HMM) is suitable to detect distant homologies (< 30%) between sequences, but performs slowly and is therefore unsuitable for large dataset screening in short time with modest computational power. As a consequence, HMM is indicated to confirm the specificity of a given *F. oxysporum* enzyme inferred by homology search with BLAST. If a sequence is considered to be specific of *Fusarium* because no significant homologous hit is found with *A. thaliana* using BLAST, but a homologous pair is detected with HMM by comparing the parasite enzyme to the proteomes of its relative plant hosts, the specificity relationship previously established with BLAST must be discarded, otherwise it can be maintained. To this end, the HMM-based lightning-fast iterative sequence search (HHblits) methodology, described by Remmert et al. (2011), was used as described in Catharina et al. (2017). The *F. oxysporum* protein sequences that are

candidates to be specific in the context of the host-parasite relation with *A. thaliana* were submitted separately to HHblits online (<http://toolkit.tuebingen.mpg.de/hhblits>) under the standard parameters proposed for a homology search using HMM. The sequences of *F. oxysporum* that hit any sequence of the fifteen plants were considered as false positive and were removed from the list of specific proteins of *F. oxysporum* for the host-parasite pair considered. Only the sequences that did not produce hit with one of the fifteen plants in the analysis were considered specific to *F. oxysporum*.

2.4. Search for false positives

By sequence homology (BLASTp), we compared the subset of specific enzymes of *F. oxysporum*, as inferred from *F. graminearum*, with the protein sequences derived from the genome annotations of *A. thaliana* and of 14 other plant species that are natural hosts of *F. oxysporum*: *B. rapa*, *Cajanus cajan*, *Carica papaya*, *G. max*, *Gossypium raimondii*, *J. curcas*, *Linum usitatissimum*, *Lotus japonicus*, *Medicago truncatula*, *Populus trichocarpa*, *Prunus persica*, *R. communis*, *Solanum lycopersicum*, *Theobroma cacao* (Table S1). These additional comparisons enabled to figure out the occurrence of false positives, but also the conservation of EC specific to *F. oxysporum* compared to a significant sample of its host range. However, we pursued the exercise of data debugging for false positives only with *A. thaliana* since it is a model and with *B. rapa*, *G. max*, *R. communis* and *J. curcas* due to their importance in biodiesel production. The procedure described by Catharina et al. (2017) was applied equally for the five plant species according to genome and protein data as available from the databases listed in Table S2.

To further straighten out our sample of putative specific enzymes, we performed homology searches (tBLASTn) between *F. oxysporum* enzymes and the DNA of plant genomes in order to find out if fungal enzyme functions might be encoded by unannotated intergenic regions.

When a hit corresponded to an annotated protein, we validate this hit by comparing the fungal enzymes (query) involved in this hit with the plant proteins produced by the gene detected by this hit (subject). Then, we performed a reverse similarity comparison (tBLASTn) of the plant protein (query) against the DNA stretch (gene coordinates plus and minus 10 kb) that produced this hit (subject) in order to confirm the gene structures (exons + introns) as available through the annotation and the consistency of its reading frame.

In cases of a hit with a non-annotated DNA region, the in-frame homologous plant sequence (subject) was searched for similarity with the fully annotated corresponding proteins in the non-redundant section (*nr*) of GenBank through the BLASTx online interface (www.blast.ncbi.nlm.nih.gov/Blast.cgi). Then, the inferred plant homolog was confirmed by back comparison to the genomic region from which it was extracted as described in more detailed previously (Catharina et al., 2017).

2.5. Metabolic pathway significance

In order to better characterize the metabolic impairment eventually caused to *F. oxysporum* by the inhibition of its specific enzymes, we manually searched their pathway centrality by locating their respective EC numbers in metabolic maps of the KEGG pathway database (<http://www.genome.jp/kegg/pathway.html>). Metabolic maps can be looked at as graphs where vertices stand for enzymes characterized by EC numbers and edges as chemical reactions. Thus, the input of a vertex in that graph is the substrate for the enzyme of this vertex and the output is its product. A metabolic route in a pathway shows how an input metabolite is transformed by the successive enzymes of that metabolic route to result in its final product. The graph is branched with parallel routes eventually converging to vertices that may articulate an entire pathway with another or a subpathway with several others in a larger graph. These converging vertices are considered to have a higher level of centrality than vertices within parallel routes because they link

subpathways between them and the metabolic flux need to go through them in order to transit between these subpathways. If a vertex separates two subpathways whose products are essential for the cell, the functional integrity of this vertex is also essential for that cell. Consequently, vertices with a higher level of centrality can be considered better molecular targets than those with a lower level of centrality because their inhibition is expected to result in a larger level of metabolic disarticulation, eventually fatal to the cell. Thus, based on the visual inspection of KEGG metabolic maps, when the enzyme considered was in a pathway without an alternative route, we considered that enzyme as central (we called it as “primary enzyme”) and suitable as a target for the development of inhibitors with reduced noxious side effects to the host. By contrast, when a specific enzyme was in a pathway with an alternative route(s), we considered it as an unsuitable target for inhibitor development (we called it as “secondary enzyme”) due to the compensation of its inhibition that would be expected by the enzymes of the alternative route(s).

3. Results

3.1. EC comparison and protein sequence retrieval

As outlined in the methodology, we considered in the first approximation that *F. graminearum* could be a suitable representative of *F. oxysporum*. Thus, when comparing *F. graminearum* to *F. oxysporum* plant hosts, we found that from the 7008 non-redundant ECs of KEGG, 879 were associated with *F. graminearum* and 1044 to *A. thaliana*, respectively.

By comparing both lists, we identified 247 EC that were specific to *F. graminearum*, from which we recovered 318 protein sequences. The BLASTp comparison of these 318 protein sequences of *F. graminearum* with the annotated proteins (proteomes) of the fifteen plant species of this study: *A. thaliana* (n = 35386), *B. rapa* (n = 41173), *C. cajan* (n = 48680), *C. papaya* (n = 27760), *G. max* (n = 73320), *G. raimondii* (n = 77267), *J. curcas* (n = 57437), *L. usitatissimum* (n = 43484), *L. japonicus* (n = 38482), *M. truncatula* (n = 62317), *P. trichocarpa* (n = 45033), *P. persica* (n = 28702), *R. communis* (n = 31221), *S. lycopersicum* (n = 34724), *T. cacao* (n = 46143) resulted in a decrease of the enzyme number that might be considered as specific to *F. graminearum* to 83, which suggests that the ECs that are specific to *F. graminearum* compared to these plant species may vary from case to case and that the number of ECs that are specific of *F. graminearum* decreases according to the increase of sample size of plant hosts considered in the BLASTp comparison.

3.2. Search for specific enzymes in *F. oxysporum*

In a first attempt, we checked for remote homology of the 83 specific proteins of *F. graminearum* with the proteome of *A. thaliana* using HHblits and we found hits for 50 of them. Therefore, these 50 *F. graminearum* enzymes were discarded from further analyzes. The remaining set contained only 33 enzymes that were apparently specific to *F. graminearum*.

At this point, we could search for specific enzymes of *F. oxysporum* rather than of *F. graminearum*. Thus, from the 33 specific enzymes of *F. graminearum* with accessions from KEGG, we retrieved the 33 corresponding accession numbers from UniProtKB. One enzyme, FG10182.1 (EC: 4.3.2.3) was updated to FG13713 in the release 83.2 (September 1st, 2017) of KEGG. When searching the orthologs of these 33 specific enzymes of *F. graminearum* in *F. oxysporum* strain Fo5176 with OrthDB, we retrieved 30 proteins (Table 1), which means that the 3 enzymes FG03537.1 (EC: 4.2.3.6), FG10343.1 (EC: 3.4.14.9) and FG11037.1 (EC: 3.2.1.151) of *F. graminearum* do not have orthologous pairs in the genome of the *F. oxysporum* strain considered. The similarity rate between the 30 specific enzymes of *F. oxysporum* and *F. graminearum* was rather high since we found that the similarity region covered, on

Table 1Enzymes considered specific to *F. graminearum* (FGR) and *F. oxysporum* strain Fo5176 (FOX) compared to *A. thaliana*, *G. max*, *B. rapa*, *R. communis*, *J. curcas*.

#	FGR acces. (KEGG)	FGR acces. ^a (UniProtKB)	FOX acces. (UniProtKB)	EC	Name	Pathway (KEGG)
1	FG02259.1	IIREZ9	F9FYW8	2.7.7.53	ATP adenyltransferase	Fgr00230
2	FG10182.1 ^b	IISA32	F9F7R0	4.3.2.3	Ureidoglycolate lyase	
3	FG10062.1	A0A098DA90	F9F7C4	1.8.4.14	L-Methionine (R)-S-oxide reductase	Fgr00270
4	FG04828.1	IIRLM2	F9FVQ5	1.13.11.52	Indoleamine 2,3-dioxygenase	Fgr00380
5	FG07058.1	IIRSE0	F9G623	1.13.11.6	3-Hydroxyanthranilate 3,4-dioxygenase	
6	FG01272.1	A0A0E0RQ52	F9F1B7	2.4.1.16	Chitin synthase	Fgr00520
7	FG10116.1	IIS097	F9FUJ3			
8	FG05066.1	A0A0E0SIE3	F9FNN9	2.1.1.17	Phosphatidylethanolamine N-methyltransferase	Fgr00564
9	FG01057.1	IIRBW6	F9FM77	3.1.4.12	Sphingomyelin phosphodiesterase 2	Fgr00600
10	FG08851.1	IIRX10	F9G156	4.2.1.79	2-Methylcitrate dehydratase	Fgr00640
11	FG08386.1	IIRVU1	F9FX57	4.4.1.17	Holocytochrome-c synthase	Fgr00860
12	FG09810.1	IIRZG9	F9FE93	3.6.1.59	DcpS - m7GpppX diphosphatase	Fgr03018
13	FG00840.1	V6QVV9	F9FPF6	2.3.1.7	Carnitine acetyltransferase	Fgr04146
14	FG01924.1	A0A098D5A0	F9FPC4			
15	FG00218.1	IIR9R3	F9G7W3	3.4.13.19	Membrane dipeptidase	NA ^d
16	FG01017.1	IIRBS7	F9F704	3.4.14.4	Dipeptidyl-peptidase III	NA
17	FG02492.1	IIRFL3	F9FX31	2.1.1.286	25S rRNA (adenine2142-N1)-methyltransferase	NA
18	FG03349.1	IIRHT5	F9FQP5	1.13.11.1	Catechol 1,2-dioxygenase	NA
19	FG03624.1	A0A119FAY2	F9FIS6	3.2.1.8	endo-1,4-Beta-xylanase	NA
20	FG10999.1	A0A119FXT9	F9FP27	3.2.1.8	endo-1,4-Beta-xylanase	NA
21	FG03990.1	IIRJH0	F9FYN8	3.1.3.84	ADP-ribose 1"-phosphate phosphatase	NA
22	FG06477.1	IIRQX6	F9G5J4	4.2.1.94	Scytalone dehydratase	NA
23	FG07016.1	IIRSA2	F9FMM7	3.2.1.101	Mannan endo-1,6-alpha-mannosidase	NA
24	FG10387.1	IIS0Z5	F9FSX4			
25	FG07974.1	IIRUS1	F9FVE8	3.2.1.40	Alpha-L-rhamnosidase	NA
26	FG09073.1	IIRXK2	F9FZ86			
27	FG08289.1	A0A0E0RZB4	F9FEN7	3.4.24.39	Deuterolysin	NA
28	FG09485.1	IIRYM8	F9FL57	5.5.1.5	Carboxy-cis,cis-muconate cyclase	NA
29	FG09999.1	IIRZY9	F9FI91	2.1.1.312	25S rRNA (uracil2843-N3)-methyltransferase	NA
30	FG11140.1 ^c	A0A098DYT2	F9FPS4	4.1.1.6	Aconitate decarboxylase	NA
31	FG03537.1	A0A119FCV5	NA	4.2.3.6	Trichodiene synthase	Fgr00909
32	FG10343.1	IIS0V3	NA	3.4.14.9	Tripeptidyl-peptidase I	NA
33	FG11037.1	IIS2N7	NA	3.2.1.151	Xyloglucan-specific endo-beta-1,4-glucanase	NA

^a Acces. is for accession.^b This accession was changed to FG13713 in KEGG release 83.2 (September 1st, 2017).^c This accession was removed in KEGG release 83.2.^d NA: No ortholog for *F. oxysporum*.

average, 97% of the query size for the homologous pairs between both fungi. Only one homologous pair had a similarity region covering 67% of the query sequence, but involved a homologous region of 368 amino acids with an identity rate of 77%. The other similarity regions between homologous pairs account for at least 80% of the query size. The average identity of the homologous pairs among the 30 specific sequences was 83.7%, with only two sequences having an identity level below 70%, i.e., 57% with a 100% coverage and 66% with a 91% coverage. These results show that we could transfer the functions of *F. graminearum* to *F. oxysporum* with an excellent level of reliability.

Although only 11 *F. oxysporum* proteins of UniProtKB had an assigned enzymatic activity, we successfully confirmed all cases of orthology by BLASTp and could, therefore, transfer their corresponding EC from KEGG when necessary. Thus, in the following, we removed FG03537.1, FG10343.1, and FG11037.1 from the analysis and considered the set of 30 enzymes outlined above as representative from the set of specific enzymes of *F. oxysporum*.

Thus, the 30 enzymes considered specific to *F. oxysporum* on the basis of its proteome comparison with those of its plant hosts were associated with 25 enzymatic activities (EC: 1.13.11.1, 1.13.11.52, 1.13.11.6, 1.8.4.14, 2.1.1.17, 2.1.1.286, 2.1.1.312, 2.3.1.7, 2.4.1.16, 2.7.7.53, 3.1.3.84, 3.1.4.12, 3.2.1.101, 3.2.1.40, 3.2.1.8, 3.4.13.19, 3.4.14.4, 3.4.24.39, 3.6.1.59, 4.1.1.6, 4.2.1.79, 4.2.1.94, 4.3.2.3, 4.4.1.17, 5.5.1.5) (Table 1). When searching these 30 enzymes for similarity (tBLASTn) with the genomes of *A. thaliana*, *B. rapa*, *G. max*, *R. communis* and *J. curcas*, we only found two enzymes whose hits were significant and could, therefore, be considered as false positives.

In one case, the comparison of F9G156 gave hits with genome regions of *A. thaliana* corresponding to AT5G65060.1 and AT5G65050.3.

When these homologous regions were compared to *nr*, they produced significant hits for the MADS-box transcription factor of *F. oxysporum* (FOSC 3-a), which suggest that F9G156 is not specific to *F. oxysporum* when compared to *A. thaliana*.

The other case was relative to F9FMM7 that gave a hit with the *J. curcas* genome corresponding to the protein jcu:105635122. When comparing (BLASTp) the homologous region with *nr*, significant hits were obtained between F9FMM7 and species of the genus *Fusarium*, which suggests that this fungal enzyme is not specific to the parasite in *J. curcas*.

3.3. Metabolic pathway significance

The ECs assigned to each enzyme of *F. graminearum* and *F. oxysporum* that successfully mapped to metabolic pathways were inserted in 11 different metabolic maps from KEGG (Table 1). These enzymes were distributed into ECs with a single route of execution and ECs with alternative routes of execution, that is, different metabolic routes for the same product. The enzymes distributed into ECs with a single route of execution that we denoted as primary enzymes were: 2.4.1.16, 3.6.1.59, 2.3.1.7, 1.13.11.6, 4.3.2.3, 4.4.1.17, 3.1.4.12, while the ECs with alternative routes of execution that we denoted secondary enzymes were: 4.2.1.79, 1.13.11.52, 2.1.1.17, 1.8.4.14, 2.7.7.53.

4. Discussion

The toxicity caused to the environment by fungicides has motivated the search of specific enzyme targets in *F. oxysporum* through the present work. Crops such as *B. rapa*, *G. max*, *J. curcas* and *R. communis* are

expected to have a large impact on the environment due to their commercial importance in the context of biodiesel production. *A. thaliana* is a model for these plant species, which has justified its choice in the context of this report.

When comparing *F. graminearum* to more than one plant host, the number of specific enzymes of this fungus tends to decrease, which shows that each host-parasite pair should be considered on its own features. Actually, such relationship is not surprising since parasites coevolved with their host in such a way to keep virulence and competition with other microbes in their environment. Thus, one expects that the genes kept in one host-parasite pair might be different from one involving a different plant species even if the fungal species itself is the same (Ma et al., 2010).

Actually, *Fusarium* spp. are haploid ascomycetous fungi with a gene number variation among species of at least 10% (Walkowiak et al., 2016). Since the plant species involved in the interaction is different, it is not surprising to find differences in numbers of genes specific to *Fusarium* spp. depending on the host-parasite system considered.

In addition, other factors such as horizontal gene transfer (HGT) can be involved, since pathogenesis (Bukhalid et al., 2002; Burger and Lang, 2003) can promote HGTs between two species (Loreto et al., 2008; Trobridge, 2009; Haegeman et al., 2011; Gao et al., 2014). Richards et al. (2009) demonstrated events of HGT occurring from *Fusarium* to rice. For instance, several plants, such as *A. thaliana*, obtained the shikimate pathway from symbiotic fungi via HGT (Richards et al., 2006; Emiliani et al., 2009; Tzin and Galili, 2010).

According to whether or not the inhibition of an enzyme may disable the metabolic pathway and bring irreversible damage to *F. oxysporum* in such a way to prevent its survival, we classified specific enzymes into key (primary) and auxiliary (secondary) ones. Primary enzymes, can be considered as promising targets for the development of new fungicides because they inhibition has the capacity to disrupt the metabolic pathway in which they are inserted and thus lead to fungal death. By contrast, secondary enzymes would have an auxiliary role because they may be inserted in alternative routes and in case of complete inhibition, they would not be able to bring the fungus to death.

4.1. Primary enzymes

Below, we give a list of key enzymes for the control of *Fusarium* spp. for the preparation of a putative cocktail of inhibitors with a brief description of their physiological roles:

- 1- In fungi, the main constituent of cell walls is the chitin polymer, which is the product of a series of biochemical reactions, where the latter consists in the synthesis of *N*-acetylglucosamine-1 phosphate by chitin synthase (EC: 2.4.1.16). Cells in which chitin formation is disrupted are affected by osmotic sensitivity, abnormal morphology, and disruption of growth (Cabib et al., 1982). Although all the enzymes of the chitin biosynthetic pathway could serve as targets in the design of antifungal molecules, most studies concern the enzyme chitin synthase (Behr, 2003). As chitin is absent in vertebrates and plants, the inhibition of its biosynthesis has been considered as a safe and selective option for the development of antifungal agents (Ruiz-Herrera and San-Blas, 2003). Therefore, the proposal was to deactivate the chitin synthase (EC: 2.4.1.16) to prevent the synthesis of chitin without damaging the plant. This idea is sustainable because chitin synthase is a fungal-specific enzyme present in an essential pathway, without an alternative route. Another possible form of inhibition of *Fusarium* growth would be through blocking the activity of the enzyme UDP-*N*-acetylglucosamine diphosphorylase (EC: 2.7.7.23), which converts *N*-acetylglucosamine-1 phosphate into UDP-GlcNAc, a monomer used by the enzyme chitin synthase (EC: 2.4.1.16) for the production of chitin by polymerization. However, UDP-*N*-acetylglucosamine diphosphorylase is also present

in plants and according to Jiang et al. (2005), Nozaki et al. (2012) and Chen et al. (2014), the synthesis of UDP-GlcNAc has a direct impact on the growth and development of plants with the consequence that the inhibition of EC: 2.7.7.23 could cause significant damages to the host. Nikkomycins and polyoxins were the first inhibitors of chitin synthase described (Chaudhary et al., 2013), but despite excellent in vitro results, nikkomycins and polyoxins showed poor in vivo activity due to inefficient transport or degradation before reaching the target enzyme (Graybill et al., 1998; Li and Rinaldi, 1999; Stevens, 2000; Behr et al., 2001; Winn et al., 2010). Ge et al. (2016) reported the synthesis and biological evaluation of 20 derivatives of 3-substituted amino-4-hydroxycoumarin as possible candidates for chitin synthase inhibitors among which a compound showed a very promising result as it produced an IC50 of 0.1 mM and, therefore, was considered the most promising for the future development of selective antifungal agents, mainly against fungi of the genus *Fusarium*.

- 2- The decapping scavenger enzyme (DcpS, m7GpppX diphosphatase, EC: 3.6.1.59) is critical in controlling mRNA stability and the regulation of gene expression (Li and Kiledjian, 2010). The enzyme is responsible for hydrolyzing the cap structure resulting from the decomposition of the 3'-5' degradation pathway (Wang and Kiledjian, 2001; Liu et al., 2002). Inhibitors for the enzyme DcpS have been described for the treatment of Spinal Muscular Atrophy (SMA). The compounds derived from C5-substituted quinazolines were able to potentially inhibit DcpS enzyme activity, with D156844 being indicated as the most potent inhibitor (Jarecki et al., 2005; Singh et al., 2008; Gopalsamy et al., 2017).
- 3- The binding of acetyl units to carnitine (Zhou and Lorenz, 2008) occurs reversibly through carnitine acetyltransferase (CATs, EC: 2.3.1.7). The transport of acetyl-CoA between intracellular compartments is mediated by carnitine acetyltransferases that reversibly bind acetyl units to the carnitine carrier molecule. Acetyl-CoA is the central metabolite that is the product and substrate of many pathways participating in carbon metabolism (Strijbis et al., 2010). During synthesis of carbon sources such as fatty acids, ethanol, or acetate, acetyl-CoA is produced at different sites in the cell, which requires the transport of acetyl units between different compartments (Kunau et al., 1995). Acetyl-CoA cannot cross organelle membranes without carnitine (Van Roermund et al., 1995). The noxious effects of CATs inhibition were described in several fungi (Bhambra et al., 2006; Franken et al., 2008; Son et al., 2012). Interestingly, Jaudzems et al. (2009) showed that the cardioprotective mildronate (3-(2,2,2-trimethylhydrazinium)-propionate) is able to inhibit CAT by binding the enzyme competitively with acetyl-CoA.
- 4- 3-Hydroxyanthranilate 3,4-dioxygenase (EC: 1.13.11.6) acts by catalyzing the opening of the ring of 3-hydroxyanthranilate in the final enzymatic step of the biosynthetic pathway of tryptophan to quinolinate. The biosynthesis of quinolinate from tryptophan is of key importance since quinolinate is the precursor of the pyridine ring of nicotinamide cofactors (Zhang et al., 2005). In addition, 3-hydroxyanthranilate 3,4-dioxygenase acts on the kynurenine pathway, which is considered the main route to generate quinolinate through the tryptophan metabolism, which justifies it as a good enzymatic target for inhibitor development (Stone and Darlington, 2002). Although the enzyme 3-hydroxyanthranilate 3,4-dioxygenase is not documented in the literature neither for *F. graminearum* nor for *F. oxysporum*, it may be considered of interest for the development of a fungicide, since it is the last enzyme involved in the biosynthesis of quinolinate and at the same time is at the entrance of the pathway for nicotinamide metabolism.
- 5- The ureidoglycolate lyase (EC: 4.3.2.3) is responsible for catalyzing the degradation of ureidoglycolate in glyoxylate and urea; it is the final step of the catabolic pathway leading from purines to urea (McIninch et al., 2003). McIninch et al. (2003) presented an inhibitor for the enzyme ureidoglycolate lyase designed to mimic the

transient species produced during its catalysis; it is called 2,4-dioxo-4-phenylbutanoic acid and is considered to be one of the most powerful competitive inhibitors with its K_i of 2.2 nM.

- 6- Holocytochrome-*c* synthase (HCCS, also known as cytochrome *c* heme lyase or simply heme lyase, EC: 4.4.1.17) catalyzes the importation of apocytochrome into the mitochondrion of animals, fungi and protozoa, through covalent attachment of heme cofactor to the cytochrome *c*, which is an essential component of the electron transport chain of mitochondria (Dumont et al., 1991; Mayer et al., 1995). The formation of holocytochrome *c* can be inhibited by deuterohemin, which is considered an analog of the substrate of HCCS (Nicholson et al., 1987).
- 7- Neutral sphingomyelinases (N-SMases, Sphingomyelin phosphodiesterase 2, EC: 3.1.4.12) are considered the main candidates to mediate the stress induced by ceramide production. Thus, N-SMase activity has been identified, characterized and cloned from bacteria, yeast as well as mammalian cells. Although the level of identity between these enzymes is low, the key residues that are involved in metal binding and catalysis are conserved (Hofmann et al., 2000). In yeast, the synthesis of sphingolipids is vital for growth and the inhibition of N-SMases induces cell death (Nagiec et al., 1997). For instance, australifungin is an inhibitor that is capable of blocking ceramide synthesis in yeast (Mandala et al., 1995), however, it is not commercially available. Nagiec et al. (1997) showed that the antifungal aureobasidin A inhibited the enzyme inositol phosphor-ylceramide synthase in yeast ($IC_{50} = 0.2$ nM).

4.2. Secondary enzymes

- 1- *Fusarium* species have genes that encode putative enzymes of the methylcitrate cycle, however, there has been no experimental evidence that the methylcitrate cycle is functional in these species (Domin et al., 2009). The 2-methylcitrate dehydratase (EC: 4.2.1.79) is responsible for the conversion of 2-methylcitrate to 2-methyl-*cis*-aconitate in the methylcitrate cycle, which is responsible for the production of highly toxic secondary metabolites during plant infection (Völkl et al., 2004; De Lucca, 2007; Jestoi, 2008; Domin et al., 2009). The accumulation of propanoyl-CoA exerts a toxic effect by inhibiting several enzymes of the primary metabolism (Brock and Buckel, 2004). To avoid this accumulation, most fungi use the methylcitrate cycle for the efficient degradation of propanoyl-CoA in pyruvate. The inhibition of this enzyme would lead to an accumulation of 2-methylcytrate; although there is an alternative route in *F. oxysporum* as can be deduced from the analysis of KEGG map fgr00640.
- 2- The indoleamine 2,3-dioxygenase (IDO, EC: 1.13.11.52) catalyzes the first step in the tryptophan (L-Trp) catabolism through the kynurenine pathway (Yuasa and Ball, 2015) and its main role is to provide nicotinamide adenine dinucleotide (NAD^+) via the kynurenine pathway (Yuasa and Ball, 2013). As an enzyme involved in human pathologies (Moon et al., 2015; Lovelace et al., 2016; Selvan et al., 2016; Brochez et al., 2017; Song et al., 2017), some inhibitors have already been described in the literature for IDO (Singh and Goel, 2016; Brochez et al., 2017; Markwalder et al., 2017). Jang et al. (2012, 2014) isolated benzomalvin E from *Penicillium* sp. FN070315 and thielavins from *Coniochaeta* sp. 10F058 and described these metabolites as inhibitors of IDO.
- 3- The phosphatidylethanolamine *N*-methyltransferase (EC: 2.1.1.17) catalyzes the transfer of one or more methyl groups from *S*-adenosyl-*L*-methionine. In fungi, the methylation pathway is used as the main route of biosynthesis of phosphatidylcholine (Kodaki and Yamashita, 1987; McGraw and Henry, 1989). Since the CDP-choline pathway is an auxiliary route to generate phosphatidylcholine, which is essential for fungi, the inhibition of this route is also required (Markham and Bainbridge, 1978). Robson et al. (1995) have demonstrated that the noncompetitive inhibitor hemicholinium-3

alters hyphae morphology in *F. graminearum* A3/5 strain, which suggests that inhibitors for this route should be considered in more details.

- 4- *L*-Methionine (*R*)-*S*-oxide reductase (EC: 1.8.4.14) catalyzes the synthesis of *L*-methionine (*L*-Met), which is an essential amino acid. Although *L*-Met synthesis is primordial for fungal growth and development, inhibition of the *L*-methionine (*R*)-*S*-oxide reductase enzyme would not be sufficient to cause the death of the pathogen. Another alternative route for the production of *L*-Met is found in *Fusarium* via the *L*-homocysteine pathway, which has two enzymes that are capable of generating *L*-Met, Homocysteine *S*-methyltransferase (EC: 2.1.1.10) and 5-methyltetrahydropteroyl-triglutamate-homocysteine-*S*-methyltransferase (EC: 2.1.1.14).
- 5- ATP adenyltransferase (EC: 2.7.7.53) is responsible for catalyzing the degradation of Ap4A into ATP and ADP in the presence of inorganic phosphate (Plateau et al., 1990). However, according to Plateau et al. (1990), the inhibition of *apa1* and/or *apa2* did not affect cell viability.

5. Conclusion

From the analysis of 318 *F. graminearum* enzymes, we retrieved 30 enzymes specific to *F. oxysporum* compared to 15 species of host plants. By comparing these 30 specific enzymes of *F. oxysporum* with the genome of *A. thaliana*, *B. rapa*, *G. max*, *J. curcas* and *R. communis*, we found 7 key enzymes whose inhibition is expected to affect significantly the fungus development and might be useful to control fusariosis. A closer look to the pathway of these key enzymes suggests that a cocktail of inhibitors could be designed against: chitin synthase, DcpS, carnitine acetyltransferase, 3-hydroxyanthranilate 3,4-dioxygenase, uridoglycolate lyase, holocytochrome synthase and sphingomyelin phosphodiesterase might allow the control of fusariosis. We also found five other specific enzymes of *Fusarium*, but they could at best only have an additive effect in the context of an inhibitory cocktail because they are inserted in alternative pathways.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gene.2018.07.003>.

Acknowledgments

This study was supported by providing a fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (<http://www.capes.gov.br/>) to LC.

Author disclosure statement

The authors declare that there are no conflicting financial interests.

References

- Agrios, G.N., 2005. Plant Pathology, 5th edn. Elsevier, Burlington, pp. 724–820.
- Anjani, K., Raouf, M.A., Desai, A.G., 2014. Evaluation of world castor (*Ricinus communis* L.) germplasm for resistance to *Fusarium* wilt (*Fusarium oxysporum* f. sp. *ricini*). Eur. J. Plant Pathol. 139, 567–578.
- Armstrong, G.M., Armstrong, J.K., 1981. *Formae speciales* and races of *Fusarium oxysporum* causing wilt diseases. In: Nelson, P.E., Toussoun, T.A., Cook, R.J. (Eds.), *Fusarium: Diseases, Biology, and Taxonomy*. Pennsylvania State University Press, University Park (457 pp.).
- Baayen, R.P., O'Donnell, K., Bonants, P.J., Cigelnik, E., Kroon, L.P., Roebroek, E.J., Waalwijk, C., 2000. Gene genealogies and AFLP analyses in the *Fusarium oxysporum* complex identify monophyletic and nonmonophyletic *formae speciales* causing wilt and rot disease. Phytopathology 90, 891–900.
- Beckman, C.H., 1987. The Nature of Wilt Diseases of Plants. APS Press, Minnesota (175 pp.).
- Behr, J., 2003. Chitin synthase as an antifungal target: recent advances. Curr. Med. Chem. Anti-Infect. Agents 2, 173–189.
- Behr, J.B., Gautier-Lefebvre, I., Mvondo-Evina, C., Guillerm, G., Ryder, N.S., 2001. Inhibition of chitin synthetase from *Saccharomyces cerevisiae* by a new UDP-GlcNAc analogue. J. Enzym. Inhib. 16, 107–112.
- Berrolal-Lobo, M., Molina, A., 2008. *Arabidopsis* defense response against *Fusarium*

- oxysporum*. Trends Plant Sci. 13, 145–150.
- Bhambra, G.K., Wang, Z.Y., Soanes, D.M., Wakley, G.E., Talbot, N.J., 2006. Peroxisomal carnitine acetyl transferase is required for elaboration of penetration hyphae during plant infection by *Magnaporthe grisea*. Mol. Microbiol. 61, 46–60.
- Brochez, L., Chevolet, I., Kruse, V., 2017. The rationale of indoleamine 2, 3-dioxygenase inhibition for cancer therapy. Eur. J. Cancer 76, 167–182.
- Brock, M., Buckel, W., 2004. On the mechanism of action of the antifungal agent propionate. Eur. J. Biochem. 271, 3227–3241.
- Bukhalid, R.A., Takeuchi, T., Labeda, D., Loria, R., 2002. Horizontal transfer of the plant virulence gene, nec1, and flanking sequences among genetically distinct *Streptomyces* strains in the Diastatochromogenes cluster. Appl. Environ. Microbiol. 68, 738–744.
- Burger, G., Lang, B.F., 2003. Parallels in genome evolution in mitochondria and bacterial symbionts. IUBMB Life 55, 205–212.
- Cabib, E., Roberts, R., Bowers, B., 1982. Synthesis of the yeast cell wall and its regulation. Annu. Rev. Biochem. 51, 763–793.
- Catharina, L., Lima, C.R., Franca, A., Guimaraes, A.C.R., Alves-Ferreira, M., Tuffery, P., Derreumaux, P., Carels, N., 2017. A computational methodology to overcome the challenges associated with the search for specific enzyme targets to develop drugs against *Leishmania major*. Bioinform. Biol. Insights 11, 1177932217712471.
- Chaudhary, P.M., Tupe, S.G., Deshpande, M.V., 2013. Chitin synthase inhibitors as antifungal agents. Mini-Rev. Med. Chem. 13, 222–236.
- Chen, Y.H., Shen, H.L., Hsu, P.J., Hwang, S.G., Cheng, W.H., 2014. N-acetylglucosamine-1-P uridylyltransferase 1 and 2 are required for gametogenesis and embryo development in *Arabidopsis thaliana*. Plant Cell Physiol. 55, 1977–1993.
- Cheng, X.X., Zhaom, L.H., Klosterman, S.J., Feng, H.J., Feng, Z.L., Wei, F., Shi, Y.Q., Li, Z.F., Zhu, H.Q., 2017. The endochitinase VDECH from *Verticillium dahliae* inhibits spore germination and activates plant defense responses. Plant Sci. 259, 12–23.
- Cooper, R.M., Wood, R.K.S., 1973. Induction of synthesis of extracellular cell-wall degrading enzymes in vascular wilt fungi. Nature 246, 309–311.
- De Lucca, A.J., 2007. Harmful fungi in both agriculture and medicine. Rev. Iberoam. Micol. 24, 3.
- de Sain, M., Rep, M., 2015. The role of pathogen-secreted proteins in fungal vascular wilt diseases. Int. J. Mol. Sci. 16, 23970–23993.
- Desmond, O.J., Manners, J.M., Stephens, A.E., Maclean, D.J., Schenk, P.M., Gardiner, D.M., Munn, A.L., Kazan, K., 2008. The *Fusarium* mycotoxin deoxynivalenol elicits hydrogen peroxide production, programmed cell death and defence responses in wheat. Mol. Plant Pathol. 9, 435–445.
- Domin, N., Wilson, D., Brock, M., 2009. Methylcitrate cycle activation during adaptation of *Fusarium solani* and *Fusarium verticillioides* to propionyl-CoA-generating carbon sources. Microbiology 155, 3903–3912.
- Dumont, M.E., Cardillo, T.S., Hayes, M.K., Sherman, F., 1991. Role of cytochrome c heme lyase in mitochondrial import and accumulation of cytochrome c in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 11, 5487–5496.
- Emiliani, G., Fondi, M., Fani, R., Gribaldo, S., 2009. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. Biol. Direct 4, 7.
- Enya, J., Togawa, M., Takeuchi, T., Yoshida, S., Tsushima, S., Arie, T., Sakai, T., 2008. Biological and phylogenetic characterization of *Fusarium oxysporum* complex, which causes yellows on *Brassica* spp., and proposal of *F. oxysporum* f. sp. *rapae*, a novel forma specialis pathogenic on *B. rapa* in Japan. Phytopathology 98, 475–483.
- Franken, J., Kroppenstedt, S., Swiegers, J.H., Bauer, F.F., 2008. Carnitine and carnitine acetyltransferases in the yeast *Saccharomyces cerevisiae*: a role for carnitine in stress protection. Curr. Genet. 53, 347.
- Gao, C., Ren, X., Mason, A.S., Liu, H., Xiao, M., Li, J., Fu, D., 2014. Horizontal gene transfer in plants. Funct. Integr. Genomics 14, 23–29.
- Ge, Z., Ji, Q., Chen, C., Liao, Q., Wu, H., Liu, X., Huang, Y., Yuan, L., Liao, F., 2016. Synthesis and biological evaluation of novel 3-substituted amino-4-hydroxycoumarin derivatives as chitin synthase inhibitors and antifungal agents. J. Enzyme Inhib. Med. Chem. 31, 219–228.
- Gopalsamy, A., Narayanan, A., Liu, S., Parikh, M.D., Kyne Jr., R.E., Fadeyi, O., Tones, M.A., Cherry, J.J., Nabhan, J.F., LaRosa, G., Petersen, D.N., Menard, C., Foley, T.L., Noell, S., Ren, Y., Loria, P.M., Maglich-Goodwin, J., Rong, H., Jones, L.H., 2017. Design of potent mRNA Decapping Scavenger Enzyme (DcpS) inhibitors with improved physicochemical properties to investigate the mechanism of therapeutic benefit in spinal muscular atrophy (SMA). J. Med. Chem. 60, 3094–3108.
- Goswami, R.S., Kistler, H.C., 2004. Heading for disaster: *Fusarium graminearum* on cereal crops. Mol. Plant Pathol. 5, 515–525.
- Graybill, J.R., Najvar, L.K., Bocanegra, R., Hector, R.F., Luther, M.F., 1998. Efficacy of nikkomycin Z in the treatment of murine histoplasmosis. Antimicrob. Agents Chemother. 42, 2371–2374.
- Gullino, M.L., Garibaldi, A., 2005. Evolution of fungal diseases of ornamental plants and their management. In: Deshmukh, S.K., Rai, M.K. (Eds.), Biodiversity of Fungi. Their Role in Human Life. Science Pub. Inc., Michigan, pp. 357–374.
- Haegeman, A., Jones, J.T., Danchin, E.G., 2011. Horizontal gene transfer in nematodes: a catalyst for plant parasitism? Mol. Plant-Microbe Interact. 24, 879–887.
- Hofmann, K., Tomiuk, S., Wolff, G., Stoffel, W., 2000. Cloning and characterization of the mammalian brain-specific, Mg²⁺-dependent neutral sphingomyelinase. Proc. Natl. Acad. Sci. 97, 5895–5900.
- Jang, J.P., Jang, J.H., Soung, N.K., Kim, H.M., Jeong, S.J., Asami, Y., Shin, K.S., Kim, M.R., Oh, H., Kim, B.Y., Ahn, J.S., 2012. Benzomalvin E, an indoleamine 2, 3-dioxygenase inhibitor isolated from *Penicillium* sp. FN070315. J. Antibiot. 65, 215.
- Jang, J.P., Jang, J.H., Oh, M., Son, S., Kim, S.M., Kim, H.M., Shin, K.S., Oh, H., Soung, N.K., Hong, Y.S., Kim, B.Y., Ahn, J.S., 2014. Inhibition of indoleamine 2, 3-dioxygenase by thielavin derivatives from a soil fungus, *Comiochaeta* sp. 10F058. J. Antibiot. 67, 331.
- Jarecki, J., Chen, X., Bernardino, A., Coovert, D.D., Whitney, M., Burghes, A., Stack, J., Pollok, B.A., 2005. Diverse small-molecule modulators of SMN expression found by high-throughput compound screening: early leads towards a therapeutic for spinal muscular atrophy. Hum. Mol. Genet. 14, 2003–2018.
- Jaudzems, K., Kuka, J., Gutsaits, A., Zinovjevs, K., Kalvinsh, I., Liepinsh, E., Liepinsh, E., Dambrova, M., 2009. Inhibition of carnitine acetyltransferase by mildronate, a regulator of energy metabolism. J. Enzyme Inhib. Med. Chem. 24, 1269–1275.
- Jestoi, M., 2008. Emerging *Fusarium*-mycotoxins fusaproliferin, beauvericin, enniatins, and moniliformin—a review. Crit. Rev. Food Sci. Nutr. 48, 21–49.
- Jiang, H., Wang, S., Dang, L., Wang, S., Chen, H., Wu, Y., Jiang, X., Wu, P., 2005. A novel short-root gene encodes a glucosamine-6-phosphate acetyltransferase required for maintaining normal root cell shape in rice. Plant Physiol. 138, 232–242.
- Jones, T.M., Anderson, A.J., Albersheim, P., 1972. Host-pathogen interactions IV. Studies on the polysaccharide-degrading enzymes secreted by *Fusarium oxysporum* f. sp. *lycopersici*. Physiol. Plant Pathol. 2, 153–166.
- Kodaki, T., Yamashita, S., 1987. Yeast phosphatidylethanolamine methylation pathway. Cloning and characterization of two distinct methyltransferase genes. J. Biol. Chem. 262, 15428–15435.
- Kumar, S., Kaushik, N., 2013. Endophytic fungi isolated from oil-seed crop *Jatropha curcas* produces oil and exhibit antifungal activity. PLoS One 8, e56202.
- Kunau, W.H., Dommers, V., Schulz, H., 1995. β -Oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress. Prog. Lipid Res. 34, 267–342.
- Li, Y., Kiledjian, M., 2010. Regulation of mRNA decapping. Wiley Interdiscip. Rev. RNA 1, 253–265.
- Li, R.K., Rinaldi, M.G., 1999. In vitro antifungal activity of nikkomycin Z in combination with fluconazole or itraconazole. Antimicrob. Agents Chemother. 43, 1401–1405.
- Lima, C.R., Carels, N., Guimaraes, A.C.R., Tuffery, P., Derreumaux, P., 2016. *In silico* structural characterization of protein targets for drug development against *Trypanosoma cruzi*. J. Mol. Model. 22, 244.
- Liu, H., Rodgers, N.D., Jiao, X., Kiledjian, M., 2002. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. EMBO J. 21, 4699–4708.
- Loreto, E.L.S., Carareto, C.M.A., Capy, P., 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity 100, 545–554.
- Lovelace, M.D., Varney, B., Sundaram, G., Franco, N.F., Ng, M.L., Pai, S., Lim, C.K., Guillemin, G.J., Brew, B.J., 2016. Current evidence for a role of the kynurenine pathway of tryptophan metabolism in multiple sclerosis. Front. Immunol. 7, 246.
- Ma, L.J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G., Diener, A., Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakhikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C., Rep, M., 2010. Comparative genomics reveals mobile pathogenicity chromosomes in the genus *Fusarium*. Nature 464, 367–373.
- Ma, L.J., Geiser, D.M., Proctor, R.H., Rooney, A.P., O'Donnell, K., Trail, F., Gardiner, D.M., Manners, J.M., Kazan, K., 2013. *Fusarium* pathogenomics. Annu. Rev. Microbiol. 67, 399–416.
- Mandala, S.M., Thornton, R.A., Frommer, B.R., Curotto, J.E., Rozdilsky, W., Kurtz, M.B., Giacobbe, R.A., Bills, G.F., Cabello, M.A., Martin, I., Peláez, F., Harris, G.H., 1995. The discovery of australifungin, a novel inhibitor of sphinganine N-acyltransferase from *Sporormiella australis*. J. Antibiot. 48, 349–356.
- Markham, P., Bainbridge, B.W., 1978. Characterization of a new choline locus in *Aspergillus nidulans* and its significance for choline metabolism. Genet. Res. 32, 303–310.
- Markwalder, J.A., Seitz, S.P., Blat, Y., Elkin, L., Hunt, J.T., Pabalan, J.G., Jure-Kunkel, M.N., Vite, G.D., Covello, K., 2017. Identification and optimization of a novel series of indoleamine 2, 3-dioxygenase inhibitors. Bioorg. Med. Chem. Lett. 27, 582–585.
- Mayer, A., Neupert, W., Lill, R., 1995. Translocation of apocytochrome c across the outer membrane of mitochondria. J. Biol. Chem. 270, 12390–12397.
- McGraw, P., Henry, S.A., 1989. Mutations in the *Saccharomyces cerevisiae* opi3 gene: effects on phospholipid methylation, growth and cross-pathway regulation of inositol synthesis. Genetics 122, 317–330.
- McIninch, J.K., McIninch, J.D., May, S.W., 2003. Catalysis, stereochemistry, and inhibition of ureidoglycolate lyase. J. Biol. Chem. 278, 50091–50100.
- Michiels, C.B., Rep, M., 2009. Pathogen profile update: *Fusarium oxysporum*. Mol. Plant Pathol. 10, 311–324.
- Moon, Y.W., Hajjar, J., Hwu, P., Naing, A., 2015. Targeting the indoleamine 2, 3-dioxygenase pathway in cancer. J. Immunother. Cancer 3, 51.
- Morrissey, J.P., Osbourn, A.E., 1999. Fungal resistance to plant antibiotics as a mechanism of pathogenesis. Microbiol. Mol. Biol. Rev. 63, 708–724.
- Nagiec, M.M., Nagiec, E.E., Baltisberger, J.A., Wells, G.B., Lester, R.L., Dickson, R.C., 1997. Sphingolipid synthesis as a target for antifungal drugs: complementation of the inositol phosphorylceramide synthase defect in a mutant strain of *Saccharomyces cerevisiae* by the AUR1 gene. J. Biol. Chem. 272, 9809–9817.
- Nelson, P.E., Dignani, M.C., Anaissie, E.J., 1994. Taxonomy, biology, and clinical aspects of *Fusarium* species. Clin. Microbiol. Rev. 7, 479–504.
- Nicholson, D.W., Köhler, H., Neupert, W., 1987. Import of cytochrome c into mitochondria. FEBS J. 164, 147–157.
- Nozaki, M., Sugiyama, M., Duan, J., Uematsu, H., Genda, T., Sato, Y., 2012. A missense mutation in the glucosamine-6-phosphate N-acetyltransferase-encoding gene causes temperature-dependent growth defects and ectopic lignin deposition in *Arabidopsis*.

- Plant Cell 24, 3366–3379.
- O'Donnell, K., Sutton, D.A., Fothergill, A., McCarthy, D., Rinaldi, M.G., Brandt, M.E., Zhang, N., Geiser, D.M., 2008. Molecular phylogenetic diversity, multilocus haplotype nomenclature, and *in vitro* antifungal resistance within the *Fusarium solani* species complex. *J. Clin. Microbiol.* 46, 2477–2490.
- Pietro, A.D., Madrid, M.P., Caracuel, Z., Delgado-Jarana, J., Roncero, M.I.G., 2003. *Fusarium oxysporum*: exploring the molecular arsenal of a vascular wilt fungus. *Mol. Plant Pathol.* 4, 315–325.
- Plateau, P., Fromant, M., Schmitter, J.M., Blanquet, S., 1990. Catabolism of bis (5'-nucleosidyl) tetraphosphates in *Saccharomyces cerevisiae*. *J. Bacteriol.* 172, 6892–6899.
- Remmert, M., Biegert, A., Hauser, A., Söding, J., 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* 9, 173–175.
- Richards, T.A., Dacks, J.B., Campbell, S.A., Blanchard, J.L., Foster, P.G., McLeod, R., Roberts, C.W., 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell* 5, 1517–1531.
- Richards, T.A., Soanes, D.M., Foster, P.G., Leonard, G., Thornton, C.R., Talbot, N.J., 2009. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21, 1897–1911.
- Robson, G.D., Wiebe, M.G., Cunliffe, B., Trinci, A.P., 1995. Choline- and acetylcholine-induced changes in the morphology of *Fusarium graminearum*: evidence for the involvement of the choline transport system and acetylcholinesterase. *Microbiology* 141, 1309–1314.
- Rost, B., 1999. Twilight zone of protein sequence alignments. *Protein Eng.* 12, 85–94.
- Ruiz-Herrera, J., San-Blas, G., 2003. Chitin synthesis as a target for antifungal drugs. *Curr. Drug Targets Infect. Disord.* 3, 77–91.
- Selvan, S.R., Dowling, J.P., Kelly, W.K., Lin, J., 2016. Indoleamine 2, 3-dioxygenase (IDO): biology and target in cancer immunotherapies. *Curr. Cancer Drug Targets* 16, 755–764.
- Singh, T., Goel, R.K., 2016. Adjuvant indoleamine 2, 3-dioxygenase enzyme inhibition for comprehensive management of epilepsy and comorbid depression. *Eur. J. Pharmacol.* 784, 111–120.
- Singh, J., Salscius, M., Liu, S.W., Staker, B.L., Mishra, R., Thurmond, J., Michaud, G., Mattoon, D.R., Printen, J., Christensen, J., Bjornsson, J.M., Pollok, B.A., Kiledjian, M., Stewart, L., Jarecki, J., Gurney, M.E., 2008. DcpS as a therapeutic target for spinal muscular atrophy. *ACS Chem. Biol.* 3, 711–722.
- Son, H., Min, K., Lee, J., Choi, G.J., Kim, J.C., Lee, Y.W., 2012. Mitochondrial carnitine-dependent acetyl coenzyme A transport is required for normal sexual and asexual development of the ascomycete *Gibberella zeae*. *Eukaryot. Cell* 11, 1143–1153.
- Song, P., Ramprasath, T., Wang, H., Zou, M.H., 2017. Abnormal kynurenine pathway of tryptophan catabolism in cardiovascular diseases. *Cell. Mol. Life Sci.* 1–18.
- Stevens, D.A., 2000. Drug interaction studies of a glucan synthase inhibitor (LY 303366) and a chitin synthase inhibitor (Nikkomycin Z) for inhibition and killing of fungal pathogens. *Antimicrob. Agents Chemother.* 44, 2547–2548.
- Stone, T.W., Darlington, L.G., 2002. Endogenous kynurenines as targets for drug discovery and development. *Nat. Rev. Drug Discov.* 1, 609–620.
- Strijbis, K., van Roermund, C.W., van den Burg, J., van den Berg, M., Hardy, G.P., Wanders, R.J., Distel, B., 2010. Contributions of carnitine acetyltransferases to intracellular acetyl unit transport in *Candida albicans*. *J. Biol. Chem.* 285, 24335–24346.
- Trobridge, G.D., 2009. Foamy virus vectors for gene transfer. *Expert. Opin. Biol. Ther.* 9, 1427–1436.
- Trusov, Y., Chakravorty, D., Botella, J.R., 2013. *Fusarium oxysporum* infection assays in *Arabidopsis*. In: *G Protein-coupled Receptor Signaling in Plants: Methods and Protocols*, pp. 67–72.
- Tzin, V., Galili, G., 2010. The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. In: *The Arabidopsis Book*, vol. 8. American Society of Plant Biologists, Washington, pp. e0132.
- Van Roermund, C.W., Elgersma, Y., Singh, N., Wanders, R.J., Tabak, H.F., 1995. The membrane of peroxisomes in *Saccharomyces cerevisiae* is impermeable to NAD (H) and acetyl-CoA under *in vivo* conditions. *EMBO J.* 14, 3480.
- Viterbo, A., Ramot, O., Chernin, L., Chet, I., 2002. Significance of lytic enzymes from *Trichoderma* spp. in the biocontrol of fungal plant pathogens. *Antonie Van Leeuwenhoek* 81, 549–556.
- Völk, A., Vogler, B., Schollenberger, M., Karlovsky, P., 2004. Microbial detoxification of mycotoxin deoxynivalenol. *J. Basic Microbiol.* 44, 147–156.
- Walkowiak, S., Rowland, O., Rodrigue, N., Subramaniam, R., 2016. Whole genome sequencing and comparative genomics of closely related *Fusarium* Head Blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genomics* 17, 1014.
- Wang, Z., Kiledjian, M., 2001. Functional link between the mammalian exosome and mRNA decapping. *Cell* 107, 751–762.
- Winn, M., Goss, R.J., Kimura, K.I., Bugg, T.D., 2010. Antimicrobial nucleoside antibiotics targeting cell wall assembly: recent advances in structure–function studies and nucleoside biosynthesis. *Nat. Prod. Rep.* 27, 279–304.
- Wolny-Koladka, K.A., 2014. *In vitro* effects of various xenobiotics on *Fusarium* spp. strains isolated from cereals. *J. Environ. Sci. Health B* 49, 864–870.
- Woloshuk, C.P., Shim, W.B., 2013. Aflatoxins, fumonisins, and trichothecenes: a convergence of knowledge. *FEMS Microbiol. Rev.* 37, 94–109.
- Yuasa, H.J., Ball, H.J., 2013. Indoleamine 2, 3-dioxygenases with very low catalytic activity are well conserved across kingdoms: IDOs of Basidiomycota. *Fungal Genet. Biol.* 56, 98–106.
- Yuasa, H.J., Ball, H.J., 2015. Efficient tryptophan-catabolizing activity is consistently conserved through evolution of TDO enzymes, but not IDO enzymes. *J. Exp. Zool. B Mol. Dev. Evol.* 324, 128–140.
- Zdobnov, E.M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R.M., Simao, F.A., Ioannidis, P., Seppely, M., Loetscher, A., Kriventseva, E.V., 2016. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* 45 (D744–D749).
- Zhang, Y., Colabroy, K.L., Begley, T.P., Ealick, S.E., 2005. Structural studies on 3-hydroxyanthranilate-3, 4-dioxygenase: the catalytic mechanism of a complex oxidation involved in NAD biosynthesis. *Biochemistry* 44, 7632–7643.
- Zhou, H., Lorenz, M.C., 2008. Carnitine acetyltransferases are required for growth on non-fermentable carbon sources but not for pathogenesis in *Candida albicans*. *Microbiology* 154, 500–509.

System biology to access target relevance in the research and development of molecular inhibitors

Larissa Catharina Costa¹, Marcio Argollo de Menezes², Nicolas Carels¹

¹Laboratório de Modelagem de Sistemas Biológicos, Centro de Desenvolvimento Tecnológico em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

²Instituto de Física, Universidade Federal Fluminense, Rio de Janeiro, Brazil

Abstract

This review focuses on how system biology may assist techniques that are used in pharmacological research, such as high throughput screening, high throughput analytical characterization of biological samples, preclinical and clinical trials, as well as targets and drug validation in order to reach patients at the lowest possible cost in a translational perspective. In signaling networks, targets can be assessed through topological criteria such as their connectivity or/and centrality. In metabolic networks, the relevance of a target for drug development may rather be assessed through some sort of enzymatic specificity resulting from remote homology, analogy or specificity in its strict sense. The concept of specificity is especially valuable in the context of a host-parasite relationship where targeting a protein specific of a parasite compared to its host is expected to minimize the noxious collateral effects of the inhibitor to the host. The relevance of putative molecular target must be proven through bench and animal validations prior to go through clinical trials. Flux balance analysis and other modeling methods of system biology enable to assess whether a molecular target can be considered as pathway's choke or not in a network context, which may facilitate the decision of developing a drug for it.

1. Introduction

The quality of pharmaceutical products started to be regulated with the inception, in the USA, of the Food and Drug Administration (FDA) in 1906. From a broad perspective, there had been a continuous growth in number of annual new therapeutic entities approved worldwide since 1940 (1). However, on a shorter time scale, the pharmaceutical industry has experienced a dramatic decrease in productivity between the 1980s and 2010 that was principally due to the cost burden of investing in the research and development (R&D) of new drugs (2). At the apogee, these costs were estimated to peak at \$1.7 billion each (3).

Far from the explosion of new drugs predicted to follow the human genome sequencing, the perceived failure of drug discovery, with an average of only two to three small-molecule drugs per year, has been attributed, in the 1990s, to the trend of pharmaceutical industry to direct its research more and more towards target-directed drug discovery. Here, by *target*, we understand polymers of amino (proteins) or nucleic (RNA) acids to which effector molecules (drug or biological inhibitor: antibody, siRNA) bind to promote their biological activity. Indeed, this genotype approach (also known as *reverse pharmacology* or *reverse chemical biology*) has been very successful when applied to well-validated targets (4). Such approach was tempting, since it allowed the rational drug development through structure-activity relationship (SAR) in the context of *one drug for*

one target, i.e., the *genes-to-drug* concept, through high throughput biochemical assays. This strategy has the advantage of testing drug candidates directly on the target, but with the drawback of neglecting possible off-target activity. Off-target(s) indicate(s) unknown secondary target(s) that positively react(s) with a compound under testing and may produce collateral effect(s). It may actually be hard to diagnose if the compound under consideration is producing effect by on-target or off-target action. Many drugs were actually discovered through their unwanted effects on off-targets rather than on the expected on-target, such as in the famous case of Viagra (5). Another drawback of the target-oriented strategies is that complex diseases usually rely on more than one target; by consequence, the inactivation of one target by drug treatment is generally null or not enough to cure patients. As a result, the activity of a rough extract on a bioassay, such as a biochemical reaction or a cell culture, often vanishes during bioguiding, i.e., the process of guiding the extraction of a compound from an extract by following its activity on a biotest at each purification step. In addition, the effect of selective interaction of a drug with a single target may be limited by network redundancies or paralogous proteins and may result in an increased risk of adaptive resistance by mutation (6).

A competitive strategy is to select compound activity based on cell phenotype (*forward pharmacology* or *classical pharmacology*) through high throughput screening (HTS), which eliminates the need of the time consuming process of suitable target identification. Through phenotype screening, one screen, in one shot, compounds that are active on a phenotype, but that also demonstrate their ability to reach their target(s) within the cell. A phenotype screening can be *up* or *down* according whether it, respectively, promotes the increase or the decrease of the readout used for the associated biotest (7). An obvious drawback of this strategy is that the precise protein targets or mechanisms of action responsible for the observed phenotypes remain to be determined, a process that is called deconvolution (8). Phenotypic screens are often optimized against mechanistic and pharmacodynamics biomarker modulation while target-based screen optimizes the activity of a drug against a given target, even if a few additional off-targets may also be affected, but without properly investigating the broader cellular activity of the agent (9).

The principle drawback of screening chemically diverse libraries with no knowledge of specific protein is the ratio of the global *in vivo* screening capacity to the total number of potentially available compounds. In contrast, chemogenomic-based drug discovery against specific targets is designed to test a large number of compounds in molecular biochemical or binding assays (10). Both biochemical and cell essays (whatever 2D, 3D as well as even 3D cell printing or ex-vivo tissues – (11–13) are simplified contexts, which fit HTS of million compounds by automated processes, but neglect the complex tissue organization and homeostasis of vertebrates. To correct such simplification, animal experimentation is necessary, however, animals are not humans and methods of target validation based on mouse genetics cannot fully predict human biology (14). There are several basic differences that may affect the transfer of results from the animal model to humans: (i) Many drugs are cross reactive to non-human primates. (ii) Treatment in animal model is frequently initiate before symptom onset, which is not the case of patients. (iii) For ethical reasons, the number of animals used is kept at minimum. (iv) Different potential side effects between animals and humans may affect the effective dosage. (v) Lack of understanding of the biology underlying the

disease. (vi) Incomplete disease correlation between animals and humans. (vii) Different drug metabolism in animals and humans in case of drug needing biological activation (15). As a consequence, a higher number of drugs interacting with less validated targets have entered clinical development during the last decade. However, this had led to a decline in the success rate of drugs during clinical development. Discrepancies between inferences drawn from animal model experimentation and clinical trials led to failures (attrition) in phase II and III due to unfavorable efficacy, lack of commercial viability and poor safety. Such compound termination late in the process of drug development has been understood as one of the main reasons of the cost burst around 2010 (1).

Another criticism that was made to pharmaceutical industry was to adhere to the *one target, one drug* paradigm hold also responsible for productivity decline (16) and that it should be replaced by a *multi-target, multi-drug* model (17), which is by definition the area of polypharmacology, i.e., the design or use of pharmaceutical agents that act on multiple targets or disease pathways. Actually, it has been suggested that even the mild inhibition of multiple molecular nodes of a cellular network can be more efficient than the complete inhibition of a single target (18), which has prompted attempts to generate global blockades of biological processes and pathways by simultaneously targeting multiple nodes in the underlying network (19). Actually, the impact of only one targeted protein on the whole pathway may be little, while the impact on that pathway will be much larger if it contains many targeted proteins and the expected impact on the whole body could be huge (20).

In response to the unsustainable situation of the pharmaceutical industry as it peaked in 2010, the process of drug R&D were shifting from how these activities were addressed as health care priorities in the past to approaches that are dominated by their potential market value. Even if the situation reverted to the figure of the 1980s (as defined by the number of new chemicals licensed by the FDA), concerns still exist regarding future decision making, which requires a new paradigm for the management of R&D activities to attend to global needs (21).

Another adaptation that occurred in response to the past experience has been to look for more stringent success criteria during the non-clinical stages in order to gain further confidence in clinical translatability. The most accepted criteria for target validation during drug discovery are based on three categories: (i) Demonstration of the target protein expression in relevant cell types or in the target tissues from animal models or patients. (ii) Demonstration that modulation of the target in cell systems results in the desired functional effect. and (iii) Demonstration that the target has a causal role in producing the disease phenotype in animal models and/or patients (22).

In addition, to effectively fight against costly termination of drugs in the clinical phase, the pharmaceutical industry has been keen to invest in theoretical and computational modeling to promote the drug discovery process (23–27), which enabled a recursive process through hypothesis testing and bench experimentation (28,29). Models are fast to execute, able to reduce the use of animals and offer cheap predictive solutions for drug pharmacokinetics (PK), pharmacodynamics (PD) as well as patient population responses.

Finally, the access to financial sources for research can be critical in some cases, such as neglected diseases, for instance, and *open science* and data sharing have received a growing interest as a mean of leveraging and combining the available resources to

accelerate drug discovery efforts (30). This community-based concept for a new drug discovery model led to the *London Declaration on Neglected Tropical Diseases* in 2012 for the control, elimination, or eradication of neglected tropical diseases (<http://unitingtocombatntds.org>).

2. Modeling strategies

Most human diseases involve sophisticated mechanistic relationships between proteins necessitating thoroughly annotated drugs by experimental means as well as the incorporation of data modeling for drug combinations (31), drug and target networks (32), and polypharmacology (33). These approaches are of particular importance in cancer and infectious disease, for which heterogeneity and evolution under the selective pressure of standard-of-care drugs result in the emergence of drug resistance.

The success of mechanism-based drug discovery relies on unambiguous evidence of the therapeutic action of drugs through clear biomolecular association, which is actually not always straightforward because of high affinity of a drug to an alternative target (off-target), for instance. This situation did evidence the need to map drugs on their protein target landscape, which led to identify protein families privileged for their historical contribution to drug discovery. In addition to privileged families, druggable protein did distribute into diverse, structurally unrelated protein families with small numbers of members mostly represented by diverse enzyme functions. Privileged families account for 44% of all human protein targets with GPCRs (12%), ion channels (19%), kinases (10%), nuclear receptors (3%) and are responsible for the therapeutic effect of 70% of small-molecule drugs (9).

As outlined above, selecting the best model organism to study a particular disease or to validate a target involves that this model organism be able to develop the disease with sufficient similarities to the human pathology under consideration. One approach for this purpose is to take the set of gene products that are modulated by current drugs and to compile the list of their orthologues in the referred model organisms. These proteins can then be analyzed for their relative affinity to their associated drugs and these drugs compared for suitability to the therapeutic indication in order to infer which therapeutic areas are potentially best mimicked by which model organism. Overall, vertebrates (dog, pig, rat, mouse and zebrafish) provide comparatively good coverage of human drug targets; as would be expected from the larger evolutionary distance and differences in anatomical systems, *Drosophila melanogaster* and *Caenorhabditis elegans* contain fewer orthologues for human disease targets (9). For pathogen targets, it is often argued that the absence of the corresponding protein in the host organism is an important prerequisite for success, and searching for proteins specific of the pathogen is often applied in bioinformatics filtering of potential targets, however, the task of identifying such suitable targets is not easy, since there are a number of proteins that are also present in humans (34).

Another challenge is how to assign targets to drugs reported to have broad mechanistic effects. The quest for these attractive multitarget medicines is progressively engaging the field of rational drug design (35) and should soon have a major role in molecular therapies. This process is assisted by several databases and algorithms that provide data on drug–target interactions with different scopes and foci. Among such tools, one may quote (i) DrugBank (36) is the most widely used and maps drugs to

proteins, which have been reported to bind to them; (ii) SuperTarget (37) is a text-mining compilation of direct and indirect drug targets; (iii) the Potential Drug Target Database ([http:// www.dddc.ac.cn/pdtd/](http://www.dddc.ac.cn/pdtd/)) that gathers known and potential drug targets with structures from the Protein Data Bank (PDB) (38), which allow the potential binding sites prediction with the structure-based druggability search engine provided by EMBL-EBI (<https://www.ebi.ac.uk/chembl/drugability>); (iv) VisANT, a network platform integrating genes, drugs, diseases, and therapies (39); (v) ChemProt, a disease chemical biology database (40); (vi) DINIES, a web interface for drug– target interaction prediction (41); (vii) VNP, a database used for visualizing the disease–target–drug interaction network (42); etc.

Large-scale integration of genomic, proteomic, signaling and metabolomics data allowed the construction of complex networks of interacting entities (proteins, small molecules, DNA elements, disease features and symptoms, etc.) by modeling different types of relationships (43), involving (i) physical protein-protein interactions (PPIs), (ii) regulatory interactions, (iii) protein-ligand interactions, (iv) natural metabolites (44), or (v) drugs (45). These networks are providing a new framework for understanding the molecular basis of physiological or pathophysiological cell states. Because of the emerging polypharmacology, it appeared that drug, target and disease spaces can be correlated and that their interrelationships can be exploited for designing drugs or cocktails, which can effectively target one or more disease states (17).

High throughput data generation has led to the development of hungry algorithms able to process massive amounts of information. Uncountable methods are already available and in development to retrieve targets and their ligands. These methods can be classified in: (i) Ligand-based approach such as pharmacophore mapping, which uses Quantitative Structure Activity Relationship (QSAR) to predict protein-ligand interaction by comparing a new ligand to the known ligands of a target protein (46,47). A pharmacophore can be identified by direct method (using receptor–ligand complexes) or by indirect method (using only a collection of ligands that are known to interact with a given receptor). The pharmacophore is then used as an abstraction of any ligand matching the target's docking side to screen large *in silico* compound datasets. The receptor-based approach for pharmacophore generation involves the mapping of physico-chemical features at the active site and their spatial relationships; an active representation of this map is then used to construct the pharmacophore model. The method takes 3D target structure as an input and a set of ligands with known activity to generate an *interaction map* that is a complement of the target's docking site. This approach is based on the assumption that the selected pharmacophore is representative for the observed activity (48). (ii) Target-based approach involving docking simulation and relying on the 3D structure of proteins to predict protein-ligand interaction (49). A more recent option is to screen drug target proteins based on primary sequence information using machines learning methods. This process inherently suffers from overfitting (50) because the training datasets have two classes. One class is called the positive dataset, which is supposed to be made of true positives, and the other is called the negative dataset, which is supposed to be made of through negatives. The accuracy and completeness of the predictions are limited by the inability to be sure that proteins in the negative dataset are indeed true negative proteins (51). However, levels of accuracy and sensitivity as high as $\geq 90\%$ were recently reached by Li et al. (2017) (52) with such machine learning methods.

(iii) Text mining that is based on keyword searching in literatures (53), but the synonymy in the name of the genes or compounds in the literatures is a major concern of this approach (54).

The integration of receptor- and ligand-based approaches through machine learning methods has led to *network pharmacology*, which is dedicated to the investigation of the complexity of polypharmacology (55). Other networks with a higher level of integration are also emerging following the same trend with, for instance, network medicine (56), diseases network (57), etc. Network pharmacology has also revealed significant correlations between drug structure similarity, target sequence similarity, and the drug–target interaction network topology, which may be used to predict unknown drug–target interaction networks from chemical structure and genomic sequence information simultaneously and on a large scale. This formalization of the drug–target interaction inference as a supervised learning problem for a bipartite graph enables the unification of the chemical and genomic in one spaces called the *pharmacological space* (58).

The combination of high-throughput experimental projects for analyzing the genome, transcriptome and proteome has allowed a better understanding of the genomic spaces populated by targets of different protein classes. In parallel, the high-throughput screening of large-scale chemical compound libraries with various biological assays enabled the exploration of the chemical space for potential drugs (59–61). The aim of chemical genomics is to relate the chemical space with the genomic one in order to identify potentially useful compounds. Actually, the number of compounds with information on their target protein is still limited, which implies that many potential interactions between chemical and genomic spaces remain to be discovered.

Effective *in silico* prediction methods are being developed to assist bench experimentation for the identification of compound–protein interactions or potential drug–target interactions that remains time consuming, costly and challenging (58). Network mining is a component of quantitative systems pharmacology that aimed at modeling the efficacy prediction of drugs addressing known or novel targets on clinical endpoints and biomarkers. Ideally, quantitative systems pharmacology is combined with disease models in a way that certain specific hypothesis are generated that can subsequently be assessed in experimental animals and finally be fed back into the *in silico* model to potentially refine the hypothesis (15).

Since network descriptions often lack the molecular details necessary to understand how the different molecular processes function, a last development has been to connect high-resolution 3D structures information in order to provide chemical means to modulate such complex systems. Actually, it has been demonstrated that a 3D model template is available for almost all of the known interactions for which there is a structure for the two monomers (62). Thus, Duran-Frigola et al., (2013) (63) combined domain-domain structural templates with a high-confidence human interactome, which allowed the rationalization of disease mutations and provide hypotheses explaining their effects at molecular level. The Interactome3D (<http://interactome3d.irbbarcelona.org>) is a resource built by these authors that provides over 12,000 protein-protein interactions (PPI) in various model organisms with structural details at atomic resolution.

As just pointed out, it is important to understand how mutations translate into the 3D structure of gene products and their consequences for drug binding because of perturbation of the binding side leading to an alteration in the treatment response.

Resources that map mutations on drug targets (64) and drug-metabolizing enzymes (65), or in a proteome-wide scale, are now emerging to include this information in the rational design of personalized therapies. Genetic variation may also cause changes in the topology of cellular networks by changing protein expression or by affecting interactions in molecular networks.

The *druggability* of a protein target is its ability to be modulated by a high affinity small molecule. There are strong evolutionary arguments why proteins have evolved molecular recognition capabilities to avoid unwanted functional disruption in the vast sea of small molecule metabolites in which they exist. Current estimates, from analysis of the pharmaceutical industry screening data, suggest that only approximately 15% of proteins expressed by an organism's genome have any inferred evidence of being potentially modulated by drug-like compounds. Additionally druggability is an attribute that is likely to be independent of lethality. Many genome-scale comprehensive knockout studies in model organisms have consistently identified around 19% of genes to be individually essential. Thus targets that are both lethal and druggable represent an intersection of less than 3% of the proteins expressed in a genome, assuming lethality and druggability are not correlated factors (10).

Here, it is important to emphasize that signal networks differ from metabolic networks by the fact that the interacting protein interfaces in PPI are difficult targets as they are usually large and flat, and often lack the cavities present at the surface of small-molecule protein receptors or enzymes (66). However, significant progresses were made in the past years (67), but there is still a need for using the knowledge of successful cases to better rationalize the chemical space (68). For instance, Koes and Camacho (2012) (69) recently developed a strategy that explores the properties of PPI interfaces to discover promising starting points for small-molecule design and several algorithms were introduced to predict PPI hot-spots (70).

3. Enzymes as targets

Enzymes are biological catalysts able to work under mild conditions of temperature and pressure that accelerate chemical reactions making them compatible with life. (71). They hold a preeminent position among protein targets because of the essentiality of their activity in many disease processes, and because the structural determinants of enzyme catalysis is suitable to inhibition by small drug-like molecules. Not surprisingly, enzyme inhibitors represent almost half the drugs in clinical use today (72). Actually, inhibitors of enzyme reactions are among the most potent and effective drugs known when their mode of action is based on competition with the original enzyme substrate (73).

Target selectivity is generally considered an important attribute for the avoidance of off-target-based toxicities (74). In preclinical compound evaluation, selectivity is most commonly measured as the ratio of drug affinity for binding to a collateral protein (off-target) to that for the target of interest (on-target) (75).

In 2002 newly launched drugs reached almost half (47%) of the drugs approved by FDA were enzymes inhibitors (76). Since that time, this proportion has diminished drastically with 70% of approved drugs in 2012 being small molecule inhibitors from which only 3% were directed toward enzymes (77).

The attractiveness of enzymes as drug targets resulted not only from the essentiality of their catalytic activity, but also from their suitability for inhibition by small molecular

weight drug-like molecules (druggability). The fact that a protein contains a druggable binding pocket does not necessarily imply that it is a good target for drug discovery; there must be some expectation that the protein plays some pathogenic role in the disease under consideration so that its inhibition will lead to a disease modification. Further the binding pocket engaged by the drug must be critical to the biological activity of the molecular target, such that interactions between the drug and the target's binding pocket lead to an attenuation of biological activity. The active sites of enzymes are usually located in surface clefts and crevices. This substrate-target interaction effectively excludes bulk solvent (water), which would otherwise reduce the catalytic activity of the enzyme. In other words, the substrate molecule is desolvated upon binding. Solvation by water is replaced by specific interactions with the protein depending on the atom arrangement in the enzyme active site, which in some way complements the structure of the substrate molecule. The inhibitor binds to the enzyme with or without competition with the substrates, but in any case leading it to decrease its catalytic activity (72).

Many proteins are potential targets for drug interventions that control human diseases. The most recent number of drug targets was estimated to be in the hundreds, based on an analysis made before 2007 (78). However, the number of druggable proteins is substantially greater according to the DrugBank database Web site (<http://www.drugbank.ca/>). The current version of this database (5.0) contains 8,206 drug entries that are linked to 4,333 nonredundant (*nr*) protein sequences (ie, drug target/enzyme/transporter/carrier).

The lack of progress, small number of drugs available in the antiparasitic class (9) and the relative toxicity of the existing ones is a motivation to explore the opportunities raised by post-genomic strategies to boost pharmaceutical research in that area.

The major categories of drug targets include enzymes, G-coupled protein receptors (GPCRs), nuclear hormone receptors, transporters, ion channels, and nucleic acids. Except for nucleic acids, all these targets elicit biological functions through ligand binding. Enzymes, however, are catalysts that make and break covalent chemical bonds, and nature has optimized enzymes for chemical transformations rather than ligand binding. This makes enzyme drug targets different, and offers opportunities for drug design that take advantage of catalysis rather than binding (73). Among target enzymes, kinases and proteases are deserving attention from pharmaceutical industry at moment because of their involvement in (i) signaling cascade (79) and (ii) multiple biological processes in all living organisms (80), respectively.

Kinases, also known as Phosphotransferases, are the most intensively studied protein drug target category in current pharmacological research, as evidenced by the vast number of kinase-targeting agents enrolled in active clinical trials (81). Kinase activity plays a key role in many cellular processes, such as cell cycle progression, apoptosis, differentiation and signal transduction (82). Eukaryotic protein kinases are related by a homologous catalytic domain of approximately 250–300 amino acids (83) and can be grouped into the serine/threonine and tyrosine kinases, which are responsible for phosphorylating the hydroxyl oxygen of their respective amino acids. Due to the pivotal role of kinases in the regulation of many cellular processes, aberrant kinase activity has been associated with a variety of diseases and the majority of human cancers (84). Aberrant kinase activity is implicated in a variety of human diseases, in particular those

involving inflammatory or proliferative responses, such as cancer, rheumatoid arthritis, cardiovascular and neurological disorders, asthma and psoriasis (85).

Proteases are encoded by more than 550 human genes (86–88) that regulate growth factors, cytokines, chemokines and cellular receptors, through activation and inactivation leading to downstream intracellular signaling and gene regulation (89). They have an important role in many signalling pathways, and represent potential drug targets for diseases ranging from cardiovascular disorders to cancer, as well as for combating many parasites and viruses (90).

Up-regulation of proteolysis is commonly associated with different types of cancer and is linked to tumor metastasis, invasion and growth (91). Dysregulated proteolysis is also a feature of various inflammatory and other, diseases.

4. System biology to predict protein targets in infectious diseases

As pointed out by Fauci and Morens (2012) (92) great pandemics and local epidemics alike have influenced the course of wars, determined the fates of nations and empires, and affected the progress of civilization, making infections compelling actors in the drama of human history'. Today, thanks to scientific achievements, pandemics are under control, however, in developed countries, pathogens are evolving toward the acquisition of antimicrobial resistance mainly because of the overuse of antibiotics in animals (92). The increasing rate of emerging and re-emerging pathogens is being observed as a consequence of the global circulation of populations from different eco-climatic conditions as well as climatic changes (93). As a matter of fact, emerging and neglected infectious diseases disproportionately afflict the poorest members of the global society. The World Health Organisation (WHO) estimated that one in six of the world's population suffers from one or more neglected infectious diseases, such as onchocerciasis, trypanosomiasis, lymphatic filariasis, schistosomiasis, soil-transmitted helminthiasis, blinding trachoma, malaria, tuberculosis and human immunodeficiency virus. Contagious diseases account for 50% of diseases in the developing countries, which represent 4.8 billion people and 80% of the world population (10). Thus, high productivity discovery systems that can be applied to a large number of pathogens are needed. The growing availability of parasite genome data provides the basis for the analysis of the pharmacological landscape of an infectious disease. However, infectious disease informatics is necessary for the rapid generation of plausible, novel medical hypotheses of testable pharmacological experiments (10).

The introduction of high-throughput sequencing or *next generation sequencing* (NGS) has revolutionized the ability to detect novel infectious agents whose genomic sequences are completely unknown or are present in extremely low numbers (94) . NGS holds the promise of identifying all potential pathogens in a single assay by comparing reads of sufficiently size to sequences of microbial genomes stored in well-annotated reference databases. Despite the deep sequence coverage of NGS, only a small fraction of short reads in clinical metagenomic samples may match pathogen sequences and such sparse reads often do not overlap sufficiently to enable *de novo* assembly into longer contiguous sequences, also called *contigs* (95). However, gene panels (96) now exist for pathogens and may be more appropriated to recover rare sequences by Ampliseq.

The most widely used approach of sample annotation in case of whole sample sequencing is to computationally subtract read contigs with the host genome sequence (e.g., human), followed by alignment to reference databases that contain sequences from candidate pathogens using the BLAST (97), or some other algorithm. Another faster strategy that may be used is to detect the coding frame of read contigs using codon bias, which is independent of the codon usage (98), and translate them into protein sequences prior to perform a BLASTp (99) or a SparkBLAST (100).

In addition, NGS and computational biology facilitate the identification of host factors that predispose or affect the response to disease. Therefore, therapies can now be tailored according to the genetic makeup of the host and the characteristics of the microbe responsible for disease. For example, *genome-wide association studies* (GWAS) enabled scientists to identify human subpopulations that have genetic variants associated with different patterns of disease progression. However, a limitation of this technique is the huge sample size necessary for statistical validation of rare mutations (94).

Pathogens must overcome host defenses and then reproduce in order to propagate, pathogen proteins interact with host proteins to either suppress or hijack the normal host protein functions. Identification of these PPIs is not only critical for understanding the biology of infection, but can also point to novel targets in treatments against human pathogens (101).

Thus, comparative signaling network investigations have also been conducted in host-parasite system and seem to be more productive in potential protein targets than enzyme based targets in terms of numbers. An example of this is the comparison of PPI in humans and *Leishmania major* that provided ~140 targets that are specific to the protozoan parasite (102).

In order to maximize the benefit to patients of therapies against infection agents, drugs should be as much as possible on-target for the infectious agents. Thus, pathogens' protein target should be (i) specific, that is, the selected target should be ideally absent from the human host because its presence could promote toxic effects from the treatment; and (ii) essential for the survival of the parasite (103,104). Thus, a drug that is able to inactivate a specific protein/enzyme of a pathogen is probably safe for its host. By contrast, a drug able to inactivate a protein/enzyme that is present in both the pathogen and the host may cause toxic effects on the host. This trivial observation leads to the concept of target specificity. One may distinguish several type of specificities in protein targets in a comparison between a parasite and its host (i) *strict specificity* when the function associated to the protein is found in the parasite, but not in its host. As a consequence, an on-target drug for that target is expected to have only few or no side effect for the host; (ii) *functional specificity*, when paralogous proteins exist in both the parasite and its host, but associated to different functions. Since both proteins are homologous, but with 3D structure optimized to interact with different substrates, a certain level of drug specificity is expected. However, the same is also true for expected collateral effects to the host since a drug could interact, even if with a different rate, with the other member of the homologous pair (34); and (iii) *analogy*, when two proteins do have the same function, but resulted from convergent evolution. Analogy is often associated to different 3D structures and may offer physico-chemical specificity suitable for drug development (105).

These concepts were used to search for specific and analogous enzymes in several host-parasite systems (106) and automated in a pipeline called AnEnPi (<http://anenpi.fiocruz.br/>). Based on the interesting source of target represented by analogous enzymes a pan-genomic investigation has been carried out in humans (107). Similar pipelines were produced independently for human vs. *Escherichia coli* host-parasite relationship (108). On a higher level of integration, several multi-omic approaches have integrated proteomics and metabolomics to obtain a systems-level understanding of metabolic pathway regulation upon infection (109,110). In these studies, the integration of protein to metabolic pathways was used to identify specific proteins that may be targeted by pathogens to cause metabolic alterations. By analyzing network topology, one can identify functional relations between nodes in the network and key regulators of a system. In an early example of multi-omic network analysis during infection, proteomic and lipidomics data during HCV infection were used to generate a network relating proteins and lipids through abundance correlations (111). As novel omics methods continue to be developed, their integration with other omics approaches will provide additional levels of information that could benefit pathogenic research.

5. Protein network

Because the number of interactions of a single protein with its neighbors may vary from one or two to dozens, a common strategy of network representation is graphs where topological features become evident at a glance. These features may, for instance, be connection hubs or cliques, modules and other regulatory motives (112).

PPIs have a key role in regulating many biological processes in metabolic, regulatory, and signaling pathways. The dysfunction of these pathways may lead to cellular diseases such as cancer and neurological disorders. Thus, network-based approaches have gained importance in drug R&D because they enable to investigate diseases as complex systems. Disease phenotypes reflect pathobiological processes that interact in a complex network (56). The high interconnectivity in disease-associated networks suggests that it is better to target entire pathways rather than single proteins, which has promoted the rise of polypharmacological approaches (113).

When inside cells, proteins typically do not function in their native state alone, but rather by interacting in concert with other proteins and metabolites, generating a high dimensional network with a complicated structure, which justifies system biology to identify essential protein targets for drug or vaccine development (102).

A PPI network involves nodes that correspond to proteins and edges, which denote interactions among proteins. High-throughput technologies allow rapid identification of PPIs and their networks (114).

PPI networks are generally considered to be scale free, and such networks typically have a few hubs (nodes with high connection degree). The hubs play a central role in a network by connecting several nodes together; thus, the inactivation of their corresponding protein, in theory, may disrupt a number of essential pathways and disarticulate a large part of the signaling network. As a result, hubs may be considered for further evaluation for being potential drug targets (115).

Nodes with high betweenness centrality could be considered as initial candidates as drug targets. Betweenness centrality characterizes the degree of influence a protein has in communicating between protein pairs and is defined as the fraction of shortest paths

going through a given node. (116). One can also employ *flux balance analysis* (FBA) of metabolic networks to find critical genes involved in the maintenance of a pathological metabolism or organism that could serve as drug targets (117).

Investigations carried out with experimental PPI networks of yeast and *Caenorhabditis elegans* have confirmed the effectiveness of topological metrics in predicting protein essentiality and demonstrated strong correlation between inferences and knockout or knockdown data (118,119). A single network covering all the genes of an organism may guide predictions at the level of individual cells or even tissues.

In yeast, Jeong et al., (2001) (118) showed the phenotypic consequence of a single gene deletion that resulted in extensive topological alterations in the complex hierarchical web of molecular interactions in a PPI network (118).

Protein connectivity in the signaling network of *C. elegans* was demonstrated to be a parameter of essentiality that is evolutionarily conserved and it has allowed the prediction of which genes are important for the majority of systematically tested phenotypes in this worm (119).

6. Metabolic networks

Advances in post-genomic era opened the door for understanding physiology from the outcome of biologically plausible constraints imposed on metabolic reaction fluxes at the molecular level. In particular, genome-scale reconstruction of metabolic networks provided valuable tools for assessing organism-wide changes resulting from point modifications of metabolic reaction fluxes both at gene and post-transcriptional levels (120).

Reconstruction of genome-scale metabolic networks containing all of the known metabolic reactions in an organism and the genes encoding each enzyme is now possible (121,122). In particular, constraint-based methods like FBA enable the prediction of phenotypic traits like optimal growth rate from the maximization of an objective function written in terms of reaction fluxes, and biologically plausible constraints (123). The most general constraint underlying FBA is mass-balance or stationarity of metabolite concentrations inside the cell, which is observed when a bacterial population doubles at regular intervals (124) and the concentration of each intracellular metabolite remains constant in time through the balance between production and consumption rates.

Cell duplication can be abstracted as the accumulation of biomass precursors in relative amounts determined by experimental assays (125). A pseudo-reaction defining growth can be added to the reconstruction as a pseudo-reaction with biomass precursor metabolites as substrates, each with (a metabolite consumption is scored as a negative value) its relative cellular amount as stoichiometric index. The flux of this reaction, f_b , can be interpreted as the growth rate and arguments based on evolution and adaptation suggest that, under favorable conditions, cells should maximize growth rates (126). Thus, reaction rates in this organism should satisfy the FBA conditions determined by formula 1:

$$\text{MAX } f_b \quad (1)$$

Given formula 2:

$$\sum_{j=1}^N S_{ij} f_j = 0 \quad \text{for all metabolites } i, \quad (2)$$

where f_j is the flux of reaction j and S_{ij} the stoichiometry of the metabolite i (with a negative or positive value if the metabolite is the substrate or product of a reaction, respectively and scored as zero if not present) in the reaction j . All reactions have upper bounds with some of which that can be experimentally determined or computationally predicted. In practice, metabolic fluxes are limited on their positive side, i.e., metabolite production, and reversible reactions are split in two parts according to the rate of conversion from substrate to product and *vice versa*.

FBA is a powerful tool for the prediction of essential reactions or essential genes from an organism. By *essential*, we mean here that the inactivation of the referred reaction or gene will break biomass production and growth rate of the modeled organism. Actually, FBA states that a gene is essential if it is constraining fluxes of associated reaction(s) to zero, which leads to a zero biomass production, i.e., $\text{MAX } f_b = 0$ (see 1), in the above definition (127–129). This statement has the corollary that critical genes can be search in a given organism in comparison to another, which is what one would expect of a molecular target in a host-parasite relationship. However, in order to be able to consider the target as potentially suitable (lead candidate), it must bear some grade of specificity of the parasite compared to its host (see discussion above). This second requirement is needed to reduce as much as possible on-target toxicity on the host. Of course, the referred specificity does not eliminate toxicity sources from off-target noxious effect on the host, but this is an issue that must be evaluated by rational drug design in relation to the parasite target. New strategies for protein targeting are now available and are commonly grouped in the category of *biopharmaceuticals* (RNAi, peptides, antibodies, etc). However, the suitability of these molecules must be analyzed case by case because they may eventually fail to reach their intracellular target by being unable to cross the outer layer of the targeted cell. Among the promising techniques, one may cite gene-silencing or post-transcriptional gene regulation by RNA interference (130), where interception and degradation of mRNA is induced by synthetic double-stranded RNA molecules (dsRNA) homologous to the target gene. Thus, in order to be considered an efficient drug, the engineered dsRNA molecules must (i) target essential genes from the pathogen, (ii) share little or no homology with the host's genes, and (iii) be able to reach its molecular target.

With the large number of entirely sequenced organisms and genome-scale reconstruction of metabolic networks, it is now possible to put host-pathogen interactions in a metabolic network perspective to predict possible organism-wide metabolic changes resulting from the underlying molecular interactions (131). We propose the combination of FBA methodology to find essential genes with similarity tests to determine which genes are specific to a pathogen. Genes which are good targets for RNA interference or any other means of molecular target inactivation (small molecule inhibitors, peptides, haptamers, antibodies) are those that are (i) critical for biomass production, in the FBA sense and (ii) specific to the pathogen's genome, found with distance measures like BLAST's e-value and score.

7. Conclusion

Information systems that could automatically search for associations between proteins, compounds and diseases are not any more a ridiculous dream. On one hand,

data are accumulating on diseases parameters such as (i) protein targets effectively druggable, (ii) group of genes or of gene products that are key in the management of a disease, (iii) profiling on individual metabolic response to drug, (iv) off-target action of drugs and their noxious collateral effect, (v) mutation frequencies in key genes of humans and parasites, and on the other hand, system biology is extending its analytical and processing tool box through the development of (i) high throughput methodologies of molecular characterization, such as cheap DNA or RNA sequencing, proteome mass spectrometry, metabolome mass spectrometry, magnetic nuclear resonance (MNR) and (ii) of automated treatments through the algorithms of bioinformatics and computational biology in the context of omics and artificial intelligence, such as machine learning, probabilistic models (Bayesian and Markovian), Boolean models, kinetic modeling through nonlinear differential equation, etc. The integration of these technologies forms the core of the system biology solution to the problem of human disease and illustrates the path that humanity is taking toward a molecular approach of precision medicine with maximum reduction of collateral effects to patients.

Essentiality and synthetic lethality (when a combination of deficiencies in the expression of two or more genes leads to cell death) can be addressed by polypharmacology to beat biochemical network wiring and dynamic rewiring (promoting escaping or compensating pathways) in pathogens as well as in cell diseases such as cancer. Druggability and selectivity are physico-chemical concepts while efficacy depends on the holistic functioning of the host-pathogen system, i.e., it extends across essential genes, synthetically lethal gene combinations, virulence factors, and other factors. To achieve the desired lethal or static effect, different degrees of inhibition may be required, which implies the necessity of tuning the type of therapeutic agents used as well as their combination and dosage.

If system biology is promising, it is also true that the labor to reach the dream outlined above is huge and has ultimately to go through clinical phase approval to be transformed in translational action. Clinical trial step is an integral part of the feedback loop in which system biology (hypotheses generation) and bench experimentation (pre-clinical validation) is engaged. The reproducibility in that experimentation cycle is an essential element that goes through what is called *good laboratory practice*, and formed by rules of conduct to respect to safely reach a given purpose. The rules form themselves through the learning curve with recommendations for NGS (132) that we give, here, as an example.

References

1. Knight-Schrijver VR, Chelliah V, Cucurull-Sanchez L, Le Novère N. The promises of quantitative systems pharmacology modelling for drug development. *Comput Struct Biotechnol J*. 2016;14:363–70.
2. Arrowsmith J. A decade of change. *Nat Rev Drug Discov*. 2012;11(1):17–8.
3. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry’s grand challenge. *Nat Rev Drug Discov*. 2010 Nov 19;
4. Butcher EC. Can cell systems biology rescue drug discovery? *Nat Rev Drug Discov*. 2005;4(6):461–7.
5. Osterloh IH. The discovery and development of of Viagra®(sildenafil citrate). *Sildenafil*. 2004;1–13.
6. Priest BT, Erdemli G. Phenotypic screening in the 21st century. *Front Pharmacol*. 2014;5:264.
7. Kaelin Jr WG. Common pitfalls in preclinical cancer target validation. *Nat Rev Cancer*. 2017;17(7):425–40.
8. Lee J, Bogyo M. Target deconvolution techniques in modern phenotypic profiling. *Curr Opin Chem Biol*. 2013;17(1):118–26.
9. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. 2017;16(1):19–34.
10. L Hopkins A, Richard Bickerton G, M Carruthers I, K Boyer S, Rubin H, P Overington J. Rapid analysis of pharmacology for infectious diseases. *Curr Top Med Chem*. 2011;11(10):1292–300.
11. Soldatow VY, LeCluyse EL, Griffith LG, Rusyn I. In vitro models for liver toxicity testing. *Toxicol Res (Camb)*. 2013;2(1):23–39.
12. Luni C, Serena E, Elvassore N. Human-on-chip for therapy development and fundamental science. *Curr Opin Biotechnol*. 2014;25:45–50.
13. Peng W, Unutmaz D, Ozbolat IT. Bioprinting towards physiologically relevant tissue models for pharmaceuticals. *Trends Biotechnol*. 2016;34(9):722–32.
14. Andrade EL, Bento AF, Cavalli J, Oliveira SK, Freitas CS, Marcon R, et al. Non-clinical studies required for new drug development-Part I: early in silico and in vitro studies, new target discovery and validation, proof of principles and robustness of animal studies. *Brazilian J Med Biol Res*. 2016;49(11):e5644.
15. Denayer T, Stöhr T, Van Roy M. Animal models in translational medicine: Validation and prediction. *New Horizons Transl Med*. 2014;2(1):5–11.
16. Wermuth CG. Multitargeted drugs: the end of the “one-target-one-disease”philosophy? *Drug Discov Today*. 2004;9(19):826–7.
17. Masoudi-Nejad A, Mousavian Z, Bozorgmehr JH. Drug-target and disease networks: polypharmacology in the post-genomic era. *silico Pharmacol*. 2013;1(1):1–17.
18. Ágoston V, Csermely P, Pongor S. Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example. *Phys Rev E*. 2005;71(5):51909.
19. Winter GE, Rix U, Carlson SM, Gleixner K V, Grebien F, Gridling M, et al. Systems-pharmacology dissection of a drug synergy in imatinib-resistant CML.

- Nat Chem Biol. 2012;8(11):905–12.
20. Chen S, Jiang H, Cao Y, Wang Y, Hu Z, Zhu Z, et al. Drug target identification using network analysis: Taking active components in Sini decoction as an example. *Sci Rep*. 2016;6:24245.
 21. Kinch MS, Haynesworth A, Kinch SL, Hoyer D. An overview of FDA-approved new molecular entities: 1827–2013. *Drug Discov Today*. 2014 Aug;19(8):1033–9.
 22. Winkler H. Target validation requirements in the pharmaceutical industry. *Targets*. 2003;2:69–71.
 23. Milligan PA, Brown MJ, Marchant B, Martin SW, Graaf PH, Benson N, et al. Model-based drug development: a rational approach to efficiently accelerate drug development. *Clin Pharmacol Ther*. 2013;93(6):502–14.
 24. Visser SAG, Manolis E, Danhof M, Kerbusch T. Modeling and simulation at the interface of nonclinical and early clinical drug development. *CPT pharmacometrics Syst Pharmacol*. 2013;2(2):1–3.
 25. Visser SAG, Aurell M, Jones RDO, Schuck VJA, Egnell A-C, Peters SA, et al. Model-based drug discovery: implementation and impact. *Drug Discov Today*. 2013;18(15):764–75.
 26. Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, et al. Lessons learned from the fate of AstraZeneca’s drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov*. 2014;13(6):419–31.
 27. Visser SAG, Alwis DP, Kerbusch T, Stone JA, Allerheiligen SRB. Implementation of quantitative and systems pharmacology in large pharma. *CPT pharmacometrics Syst Pharmacol*. 2014;3(10):1–10.
 28. Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Applying computational modeling to drug discovery and development. *Drug Discov Today*. 2006;11(17):806–11.
 29. Brodland GW. How computational models can help unlock biological systems. *Semin Cell Dev Biol*. 2015;47:62–73.
 30. Bombelles T, Coaker H. Neglected tropical disease research: rethinking the drug discovery model. *Future Med Chem*. 2015;7(6):693–700.
 31. Al-Lazikani B, Workman P. Unpicking the combination lock for mutant BRAF and RAS melanomas. *Cancer Discov*. 2013;3(1):14–9.
 32. Workman P, Clarke PA, Al-Lazikani B. Blocking the survival of the nastiest by HSP90 inhibition. *Oncotarget*. 2016;7(4):3658.
 33. Paolini G V, Shapland RHB, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. *Nat Biotechnol*. 2006;24(7):805–15.
 34. Catharina L, Carlyle Ribeiro Lima AF, Guimarães ACR, Alves-Ferreira M, Tuffery P, Derreumaux P, et al. A Computational Methodology to Overcome the Challenges Associated With the Search for Specific Enzyme Targets to Develop Drugs Against. *Bioinform Biol Insights*. 2017;11:1–21.
 35. Wei D, Jiang X, Zhou L, Chen J, Chen Z, He C, et al. Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching. *J Med Chem*. 2008;51(24):7882–8.
 36. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34(suppl_1):D668–72.

37. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, et al. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* 2007;36(suppl_1):D919–22.
38. Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, et al. PDTD: a web-accessible protein database for drug target identification. *BMC Bioinformatics.* 2008;9(1):104.
39. Hu Z, Chang Y-C, Wang Y, Huang C-L, Liu Y, Tian F, et al. VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res.* 2013;41(W1):W225–31.
40. Kim Kjørulff S, Wich L, Kringelum J, Jacobsen UP, Kouskoumvekaki I, Audouze K, et al. ChemProt-2.0: visual navigation in a disease chemical biology database. *Nucleic Acids Res.* 2012;41(D1):D464–9.
41. Yamanishi Y, Kotera M, Moriya Y, Sawada R, Kanehisa M, Goto S. DINIES: drug–target interaction network inference engine based on supervised analysis. *Nucleic Acids Res.* 2014;42(W1):W39–45.
42. Hu Q, Deng Z, Tu W, Yang X, Meng Z, Deng Z, et al. VNP: interactive visual network pharmacology of diseases, targets, and drugs. *CPT pharmacometrics Syst Pharmacol.* 2014;3(3):1–8.
43. Barabasi A-L, Oltvai ZN. Network biology: understanding the cell’s functional organization. *Nat Rev Genet.* 2004;5(2):101–13.
44. Yamada T, Bork P. Evolution of biomolecular networks—lessons from metabolic and protein interactions. *Nat Rev Mol Cell Biol.* 2009;10(11):791–803.
45. Yildirim MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M. Drug—target network. *Nat Biotechnol.* 2007;25(10):1119–26.
46. Butina D, Segall MD, Frankcombe K. Predicting ADME properties in silico: methods and models. *Drug Discov Today.* 2002;7(11):S83–8.
47. Byvatov E, Fechner U, Sadowski J, Schneider G. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci.* 2003;43(6):1882–9.
48. Khedkar SA, Malde AK, Coutinho EC, Srivastava S. Pharmacophore modeling in drug discovery and development: an overview. *Med Chem (Los Angeles).* 2007;3(2):187–97.
49. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009;30(16):2785–91.
50. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, et al. DrugBank 3.0: a comprehensive resource for “omics” research on drugs. *Nucleic Acids Res.* 2010;39(suppl_1):D1035–41.
51. Wang JT, Liu W, Tang H, Xie H. Screening drug target proteins based on sequence information. *J Biomed Inform.* 2014;49:269–74.
52. Li Z, Han P, You Z-H, Li X, Zhang Y, Yu H, et al. In silico prediction of drug-target interaction networks based on drug chemical structure and protein sequences. *Sci Rep.* 2017;7(1):11174.
53. Lam MPY, Venkatraman V, Xing Y, Lau E, Cao Q, Ng DCM, et al. Data-driven approach to determine popular proteins for targeted proteomics translation of six organ systems. *J Proteome Res.* 2016;15(11):4126–34.

54. Zhu S, Okuno Y, Tsujimoto G, Mamitsuka H. A probabilistic model for mining implicit “chemical compound–gene”relations from literature. *Bioinformatics*. 2005;21(suppl_2):ii245-ii251.
55. Wang Z, Li J, Dang R, Liang L, Lin J. PhIN: a protein pharmacology interaction network database. *CPT pharmacometrics Syst Pharmacol*. 2015;4(3):160–6.
56. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12(1):56–68.
57. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. *Proc Natl Acad Sci*. 2007;104(21):8685–90.
58. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*. 2008;24(13):i232–40.
59. Stockwell BR. Chemical genetics: ligand-based discovery of gene function. *Nat Rev Genet*. 2000;1(2):116–25.
60. Dobson CM. Chemical space and biology. *Nature*. 2004;432(7019):824–8.
61. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. 2006;34:D354–7.
62. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc Natl Acad Sci*. 2012;109(24):9438–41.
63. Duran-Frigola M, Mosca R, Aloy P. Structural systems pharmacology: the role of 3D structures in next-generation drug development. *Chem Biol*. 2013;20(5):674–84.
64. Yang JO, Oh S, Ko G, Park S-J, Kim W-Y, Lee B, et al. VnD: a structure-centric database of disease-related SNPs and drugs. *Nucleic Acids Res*. 2010;39(suppl_1):D939–44.
65. Preissner S, Kröll K, Dunkel M, Senger C, Goldsobel G, Kuzman D, et al. SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. *Nucleic Acids Res*. 2009;38(suppl_1):D237–43.
66. Fuller JC, Burgoyne NJ, Jackson RM. Predicting druggable binding sites at the protein–protein interface. *Drug Discov Today*. 2009;14(3):155–61.
67. Schlecht U, Miranda M, Suresh S, Davis RW, Onge RPS. Multiplex assay for condition-dependent changes in protein–protein interactions. *Proc Natl Acad Sci*. 2012;109(23):9213–8.
68. Sperandio O, Reynès CH, Camproux A-C, Villoutreix BO. Rationalizing the chemical space of protein–protein interaction inhibitors. *Drug Discov Today*. 2010;15(5):220–9.
69. Koes DR, Camacho CJ. PocketQuery: protein–protein interaction inhibitor starting points from protein–protein interaction structure. *Nucleic Acids Res*. 2012;40(W1):W387–92.
70. Taboureau O, Baell JB, Fernández-Recio J, Villoutreix BO. Established and emerging trends in computational drug discovery in the structural genomics era. *Chem Biol*. 2012;19(1):29–41.
71. Świderek K, Tuñón I, Moliner V, Bertran J. Computational strategies for the design

- of new enzymatic functions. *Arch Biochem Biophys*. 2015;582:68–79.
72. Copeland RA. Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists. John Wiley & Sons; 2013.
 73. Robertson JG. Enzymes as a special class of therapeutic target: clinical drugs and modes of action. *Curr Opin Struct Biol*. 2007;17(6):674–9.
 74. Copeland RA. The dynamics of drug-target interactions: drug-target residence time and its impact on efficacy and safety. *Expert Opin Drug Discov*. 2010;5(4):305–10.
 75. Copeland RA. Evaluation of enzyme inhibitors in drug discovery. A guide for medicinal chemists and pharmacologists. *Methods Biochem Anal*. 2005;46:1–265.
 76. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002;1(9):727–30.
 77. Thomas D. A Big Year for Novel Drugs Approvals [Internet]. 2013 [cited 2017 Jan 1]. Available from: <http://www.biotech-now.org/business-and-investments/inside-bio-ia/2013/01/a-big-year-for-novel-drugs-approvals>
 78. Bull SC, Doig AJ. Properties of Protein Drug Target Classes. *PLoS One*. 2015;10(3):e0117955.
 79. Burkhard K, Shapiro P. Use of inhibitors in the study of MAP kinases. *MAP Kinase Signal Protoc Second Ed*. 2010;107–22.
 80. López-Otín C, Bond JS. Proteases: multifunctional enzymes in life and disease. *J Biol Chem*. 2008;283(45):30433–7.
 81. Rask-Andersen M, Zhang J, Fabbro D, Schiöth HB. Advances in kinase targeting: current clinical use and clinical trials. *Trends Pharmacol Sci*. 2014;35(11):604–20.
 82. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science (80-)*. 2002;298(5600):1912–34.
 83. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J*. 1995;9(8):576–96.
 84. Engh RA, Bossemeyer D. Structural aspects of protein kinase control—role of conformational flexibility. *Pharmacol Ther*. 2002;93(2):99–111.
 85. Melnikova I, Golden J. Targeting protein kinases. *Nat Rev Drug Discov*. 2004;3(12):993–4.
 86. Puente XS, Sánchez LM, Overall CM, López-Otín C. Human and mouse proteases: a comparative genomic approach. *Nat Rev Genet*. 2003;4(7):544–58.
 87. Puente XS, López-Otín C. A genomic analysis of rat proteases and protease inhibitors. *Genome Res*. 2004;14(4):609–22.
 88. Puente XS, Sánchez LM, Gutiérrez-Fernández A, Velasco G, López-Otín C. A genomic view of the complexity of mammalian proteolytic systems. *Biochem Soc Trans*. 2005 Apr 1;33(2):331–4.
 89. Rodríguez D, Morrison CJ, Overall CM. Matrix metalloproteinases: what do they not do? New substrates and biological roles identified by murine models and proteomics. *Biochim Biophys Acta (BBA)-Molecular Cell Res*. 2010;1803(1):39–54.
 90. Drag M, Salvesen GS. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov*. 2010;9(9):690–701.
 91. Duffy MJ, McGowan PM, Gallagher WM. Cancer invasion and metastasis: changing views. *J Pathol*. 2008;214(3):283–93.
 92. Fauci AS, Morens DM. The perpetual challenge of infectious diseases. *N Engl J*

- Med. 2012;366(5):454–61.
93. Fontana JM, Alexander E, Salvatore M. Translational research in infectious disease: current paradigms and challenges ahead. *Transl Res.* 2012;159(6):430–53.
 94. Wooller SK, Benstead-Hume G, Chen X, Ali Y, Pearl FMG. Bioinformatics in translational drug discovery. *Biosci Rep.* 2017 Aug 31;37(4):BSR20160180.
 95. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* 2014;24(7):1180–92.
 96. Frey KG, Bishop-Lilly KA. Next-Generation Sequencing for Pathogen Detection and Identification. *Methods Microbiol.* 2015;42:525–54.
 97. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
 98. Carels N, Frías D. A Statistical Method without Training Step for the Classification of Coding Frame in Transcriptome Sequences. *Bioinform Biol Insights.* 2013 Jan 9;7:35–54.
 99. Carels N, Gumiel M, da Mota FF, de Carvalho Moreira CJ, Azambuja P. A Metagenomic Analysis of Bacterial Microbiota in the Digestive Tract of Triatomines. *Bioinform Biol Insights.* 2017;11:1177932217733422.
 100. de Castro MR, dos Santos Tostes C, Dávila AMR, Senger H, da Silva FAB. SparkBLAST: scalable BLAST processing using in-memory operations. *BMC Bioinformatics.* 2017;18(1):318.
 101. Beltran PMJ, Federspiel JD, Sheng X, Cristea IM. Proteomics and integrative omic approaches for understanding host–pathogen interactions and infectious diseases. *Mol Syst Biol.* 2017;13(3):922.
 102. Flórez AF, Park D, Bhak J, Kim B-C, Kuchinsky A, Morris JH, et al. Protein network prediction and topological analysis in *Leishmania major* as a tool for drug target selection. *BMC Bioinformatics.* 2010;11(1):484.
 103. Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee PW-N. Metabolic control analysis in drug discovery and disease. *Nat Biotechnol.* 2002;20(3):243–9.
 104. Haanstra JR, Gerding A, Dolga AM, Sorgdrager FJH, Buist-Homan M, Du Toit F, et al. Targeting pathogen metabolism without collateral damage to the host. *Sci Rep.* 2017;7:40406.
 105. Capriles PVSZ, Baptista LPR, Guedes IA, Guimarães ACR, Custódio FL, Alves-Ferreira M, et al. Structural modeling and docking studies of ribose 5-phosphate isomerase from *Leishmania major* and *Homo sapiens*: A comparative analysis for Leishmaniasis treatment. *J Mol Graph Model.* 2015 Feb;55:134–47.
 106. Otto TD, Guimarães ACR, Degraeve WM, de Miranda AB. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics.* 2008;9:544.
 107. Piergiorgio RM, de Miranda AB, Guimarães AC, Catanho M. Functional Analogy in Human Metabolism: Enzymes with Different Biological Roles or Functional Redundancy? *Genome Biol Evol.* 2017;9(6):1624–36.
 108. Mondal SI, Ferdous S, Jewel NA, Akter A, Mahmud Z, Islam MM, et al. Identification of potential drug targets by subtractive genome analysis of *Escherichia coli* O157: H7: an in silico approach. *Adv Appl Bioinforma Chem*

- AABC. 2015;8:49.
109. Su L, Zhou R, Liu C, Wen B, Xiao K, Kong W, et al. Urinary proteomics analysis for sepsis biomarkers with iTRAQ labeling and two-dimensional liquid chromatography–tandem mass spectrometry. *J Trauma Acute Care Surg*. 2013;74(3):940–5.
 110. Villar M, Ayllón N, Alberdi P, Moreno A, Moreno M, Tobes R, et al. Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by *Anaplasma phagocytophilum* infection in tick cells. *Mol Cell Proteomics*. 2015;14(12):3154–72.
 111. Diamond DL, Syder AJ, Jacobs JM, Sorensen CM, Walters K-A, Proll SC, et al. Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. *PLoS Pathog*. 2010;6(1):e1000719.
 112. Salazar GA, Meintjes A, Mazandu GK, Rapanoël HA, Akinola RO, Mulder NJ. A web-based protein interaction network visualizer. *BMC Bioinformatics*. 2014;15(1):129.
 113. Pujol A, Mosca R, Farrés J, Aloy P. Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol Sci*. 2010;31(3):115–23.
 114. Hormozdiari F, Salari R, Bafna V, Sahinalp SC. Protein-protein interaction network evaluation for identifying potential drug targets. *J Comput Biol*. 2010;17(5):669–84.
 115. Holme P, Kim BJ, Yoon CN, Han SK. Attack vulnerability of complex networks. *Phys Rev E*. 2002;65(5):56109.
 116. Estrada E. Protein Bipartivity and Essentiality in the Yeast Protein– Protein Interaction Network. *J Proteome Res*. 2006;5(9):2177–84.
 117. Kell DB. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov Today*. 2006;11(23):1085–92.
 118. Jeong H, Mason SP, Barabási A-L, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001;411(6833):41–2.
 119. Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM. A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nat Genet*. 2008;40(2):181–8.
 120. Kaldorf M, Srivastava M, Gupta SK, Liang C, Binder J, Dietl A-M, et al. Systematic Identification of Anti-Fungal Drug Targets by a Metabolic Network Approach. *Front Mol Biosci*. 2016;3:22.
 121. Joyce AR, Palsson BØ. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol cell Biol*. 2006;7(3):198–210.
 122. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*. 2009;5(1):320.
 123. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010;28(3):245–8.
 124. Neidhardt FC. Bacterial Growth: Constant Obsession with dN/dt. *J Bacteriol*. 1999;181(24):7405–8.
 125. Feist AM, Palsson BO. The biomass objective function. *Curr Opin Microbiol*. 2010;13(3):344–9.
 126. Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive

- evolution to achieve in silico predicted optimal growth. *Nature*. 2002;420(6912):186–9.
127. Joyce AR, Palsson BØ. Predicting gene essentiality using genome-scale in silico models. *Microb Gene Essentiality Protoc Bioinforma*. 2008;433–57.
 128. Sylke M. *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery*. Vol. 7. John Wiley & Sons; 2016. 1-576 p.
 129. Tobalina L, Pey J, Rezola A, Planes FJ. Assessment of FBA Based Gene Essentiality Analysis in Cancer with a Fast Context-Specific Network Reconstruction Method. *PLoS One*. 2016;11(5):e0154583.
 130. Koch A, Biedenkopf D, Furch A, Weber L, Rossbach O, Abdellatef E, et al. An RNAi-based control of *Fusarium graminearum* infections through spraying of long dsRNAs involves a plant passage and is controlled by the fungal silencing machinery. *PLoS Pathog*. 2016;12(10):e1005901.
 131. Peyraud R, Dubiella U, Barbacci A, Genin S, Raffaele S, Roby D. Advances on plant–pathogen interactions from molecular toward systems biology perspectives. *Plant J*. 2017;90(4):720–37.
 132. Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, et al. College of American Pathologists’ laboratory standards for next-generation sequencing clinical tests. *Arch Pathol Lab Med*. 2014;139(4):481–93.